

GRGETNTGADKSKSLADGMSLLGAIYILHREIEVARQVILRLFGTLLDAAPTLAGL
 DWKTSLOBLTARGMGVPSYVVTSTGPDHDKFTAVVWVDTGVSGIGHSKKEAEQK
 AASAAWKALDVLGGVGTSTV
 complement(1980. .2177)
 /gene="rnc"
 /note="Pfam match to entry PF00035 dsm, Double-stranded
 RNA binding motif, score 84.20, E-value 2.7e-21"
 complement(2259. .2540)
 /gene="rnc"
 /note="Pfam match to entry PF00636 Ribonuclease_3, RNase3
 domain, score 155.60, E-value 8.2e-43"
 complement(2514. .2540)
 /gene="rnc"
 /note="PS00517 Ribonuclease III family signature"
 complement(2657. .3310)
 /gene="ML1660"
 complement(2657. .3310)
 /gene="ML1660"
 /note="Similar to several proteins of undefined function
 e.g. Mycobacterium tuberculosis RV2926C TR:YT26_MYCTU."
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="CAC30613.1"
 /db_xref="GI:13093431"
 /db_xref="SWISS-PROT:O69468"
 /translation="MTVALALRDCRISTMTQYGVTTQRLHLSPPVFDITPLGRRP
 GAITALQRTVEISLIGLELVIEMGAPINLDRVESVEDVLVAGTVTPTVSECVR
 CLTAVHGVVTLQGLFAYPSATKVTEDDAVGHVVDGTDLEQSIIDAVGLELPPA
 PMCRSDCPGLCAGCGTSLVVEPQPHDRIDPWAALDMLAPVQPTSETDGRSEW"
 complement(3344. .4087)
 /gene="ML1661"
 complement(3344. .4087)
 /gene="ML1661"
 /note="Similar to several proteins of undefined function
 e.g. Mycobacterium tuberculosis RV2927C TR:YT27_MYCTU."
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="CAC30614.1"
 /db_xref="GI:13093432"
 /db_xref="SWISS-PROT:Q9CB86"
 /translation="MAVVRPEALDELGAIVBERAGVPMTAGCVPRGDVLELIDDIK
 DAIPGELDDQDLVDADAMLDNAKADMSVSAFTSESLSHARAEADRLISAK
 SQVDRMASEARQHSERVLGDAIESIRIATVAKREVSASRAQSECDRIENGSLY
 EKAIQGKIQQRULVSNEVVQANAESTRLIDTAHAADRLGRCODIYVDNKLAFBE
 EFLNGTISVGRGHQRTAAGTHDYTR"
 complement(4030. .4095)
 /gene="ML1662"
 /pseudo
 4317. .4897
 /gene="ML1662"
 /note="Possible pseudogene of M. tuberculosis orthologue
 RV2963 (Best blastx score 136)"
 /pseudo
 /codon_start=1
 /transl_table=11
 /product="integral membrane protein (pseudogene)"
 4905. .6083
 /note="1179 bp 100% identical to region 545582. .544404"
 complement(6182. .6655)
 /gene="coad"
 /note="synonym: ML1663"
 complement(6182. .6655)
 /gene="coad"
 /note="Highly similar to many phosphopantetheine
 adenylyltransferases (EC 2.7.7.3) involved in coenzyme A
 biosynthesis, including: Escherichia coli SW:COAD_ECOLI
 (P23875) (159 aa); Fasta score E(): 5.9e-25, 43.5%
 identity in 154 aa overlap and Mycobacterium tuberculosis
 RV2965C SW:COAD_MYCTU (Q50452) (161 aa); Fasta score E():
 0. 85.6% identity in 153 aa overlap. Contains Pfam match

to entry PF01467 Cytidylyltransf, Cytidylyltransferase."
 /codon_start=1
 /transl_table=11
 /product="probable phosphopantetheine adenylyltransferase"
 /protein_id="CAC30616.1"
 /db_xref="GI:13093433"
 /db_xref="GOA:O69466"
 /translation="WVCPGSPFPTLGHIDVFERAAQDFEVVAILINPVKGMFPL
 DERAMINSTEHLNLRVEAGELVVALVRSGMTAIVKGLRTGVDFEYELQMAQMN
 KHIAGVDITFVATAPRYSFVSSLVKVEAMLGDSVSELLPESVNRFRKNSGTS"
 complement(6257. .6652)
 /gene="coad"
 /note="Pfam match to entry PF01467 Cytidylyltransf,
 Cytidylyltransferase, score 157.90, E-value 1.7e-43"
 complement(6687. .7160)
 /gene="ML1664"
 complement(6687. .7160)
 /gene="ML1664"
 /note="Similar to several proteins of undefined function
 including: Neisseria meningitidis possible DNA methylase
 TR:CAE83466 (EMBL:AL162752) (189 aa); Fasta score E():
 3.6e-07, 39.6% identity in 96 aa overlap and Mycobacterium
 tuberculosis RV2966C TR:P95128 (EMBL:Z83018) (188 aa);
 Fasta score E(): 0. 72.9% identity in 155 aa overlap.
 Contains PS00092 N-6 Adenine-specific DNA methylases
 signature."
 /codon_start=1

Query Match 92.0%; Score 18.4; DB 1; Length 348950;
 Best Local Similarity 95.0%; Pred. No. 4.4e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCCGACACCTCGACTT 20
 |||||
 DB 1854 GCGCCGACACCTCTACTT 1873

RESULT 14
 SNO224512/c
 LOCUS
 DEFINITION Streptomyces nogalater nogalamycin biosynthesis genes.
 ACCESSION AJ224512 248262
 VERSION AJ224512.2 GI:6093257
 KEYWORDS 2,3-dehydratase; 3,5-epimerase; activator; aklanonic acid;
 aromatase; dtp-4-keto-6-deoxyhexose reductase; epimerase;
 glycosyltransferase; methylase; N-methylase; nogalamycin resistance;
 O-methylase; oxygenase; polyketide; polyketide ketoreductase; snoa1
 gene; Snoa1 protein; snoa2 gene; Snoa2 protein; snoa3 gene; Snoa3
 protein; snoaB gene; SnoaB protein; snoaC gene; SnoaC protein;
 snoad gene; Snoad protein; snoas gene; Snoas protein; snogf gene;
 snogf protein; snogg gene; snogg protein; snogh gene; snogh
 protein; snogX gene; snogX protein; snogY gene; snogY protein;
 snogZ gene; snogZ protein; snop gene; snop protein; snorA gene;
 snorA protein; snoro gene; snoro protein.
 Streptomyces nogalater
 Streptomyces nogalater
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
 Ylthonko, K., Tuikkanen, J., Jussila, S., Cong, L. and Mantsala, P.
 A gene cluster involved in nogalamycin biosynthesis from
 Streptomyces nogalater: sequence analysis and complementation of
 early-block mutations in the anthracycline pathway
 Mol. Gen. Genet. 251 (2), 113-120 (1996)
 96242142
 MEDLINE 8668120
 PUBMED 8668120

REFERENCE 2
 Ylthonko, K., Fakala, J., Kunnari, T. and Mantsala, P.
 Production of hybrid anthracycline antibiotics by heterologous
 expression of Streptomyces nogalater nogalamycin biosynthesis genes
 Microbiology (Reading, Engl.) 142 (Pt 8), 1965-1972 (1996)
 96349102
 MEDLINE

misc_feature

misc_feature

misc_feature

gene

CDS

gene

CDS

RBS

gene

CDS

repeat_unit

gene

CDS


```

WNRFPWLPAGRFIGSHLSERKSGKAGGMWTLKGGANGSGPRRRGGGGGAGGERR
REGIFARARKGGRGF"
/ gene
complement (3136..4110)
/ gene="OJ1341_A08.104"
complement (3136..4110)
/ note="start and end point are not identified"
/ gene="OJ1341_A08.104"
complement (3136..4110)
/ gene="OJ1341_A08.104"
/ note="predicted by GeneMark.hmm etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC22532.1"
/ db_xref="GI:24417171"
/ translation="MADGLAEVAPREHPRGAAGQGHVDGPDRLGRS IERPRAGAAROR
GRAATGEGGSRGSGRDPGRQWAAALDVAHTGSGGEGRSGPRGRRGHV
GKAADDERWRFSGRSGRSGRPRRCRGGQDLGSGKEGLEGEQRASRHGEHQ
PSTFVRLTGETITVAPGGEVVAEGALVLAIPREATAVREYVGNCGNQSIAANV
ATSDRWKWCSGGGLGKSGRPGSCDGEVNGAGSKSRWLKRWREATEANGGGGSG
TRCGGSDRDTGRGSRGVRGADGGSGTARRPWRREAAAGARR"
5036..10775
/ gene="OJ1341_A08.105"
join (<5036..5137,5249..5378,5898..5969,6412..7438,
9389..9523,10339..10465,10638..10775)
/ gene="OJ1341_A08.105"
/ note="start and end point are not identified"
join (5036..5137,5249..5378,5898..5969,6412..7438,
9389..9523,10339..10465,10638..10775)
/ gene="OJ1341_A08.105"
/ note="contains EST(s): AU078130(E20319)"
/ codon_start=1
/ product="unknown protein"
/ protein_id="BAC83014.1"
/ db_xref="GI:34393701"
/ translation="MERDPCDGLCRFOIHLGTGDCSLPAGGKVKPVQLCTSGQRN
YSNARGYGRSTRTVLSLNTKSTOCTVHMVPIAGVLIRPERLGLVFTAAAGT
LCYDTPMLPVIPEYVYAIELRVAALPRCHLSTPPLVVLVLRGALLSWATGP
ADFSPLHRGVAVLSSTAASOLASSSPFAHQPRPHWLSAIQCLPLLRASPP
VLCVATVLAALGWSVLYOMTDIAVQAGSATSPSSSMTHRRORRIFLDYSLFSG
NCVLQQLSLYALAPRSPWPSLLVSDIGVFWVPALEVYQHERTHSSRVPG
SDKPCVTGRPSDFDYGSSASIASIPSSARAPRAPHSTAPRHSVAVRLIH
QLPFDGYDHYGLDHSGLTPYALATSTTAQARIRIEHSCRLFLQSNWTSIS
SHWLLRQKRCPLHTASSNSSLNCSQSORQRCXDLCLANNSSDDHTD
IETGNEVNYGTTTDLDAETSYDNHISFYVAVLHVNPHITAFDGLPFPKNLHVAL
FISEKFPQSM"
complement (11285..13284)
/ gene="OJ1341_A08.106"
complement (11285..13284)
/ note="probably inactive due to including stop codon(s) in
CDS
pseudogene, gag/pol polyprotein"
/ pseudo
complement (14421..14678)
/ gene="OJ1341_A08.107"
complement (14421..14678)
/ note="start and end point are not identified"
complement (14421..14678)
/ gene="OJ1341_A08.107"
/ note="predicted by GlimmerM etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83015.1"
/ db_xref="GI:34393702"
/ translation="MGLFRRVIOATLFHAYLQOQVKPTRHPKRLKWKQGGSRHLHD
DVNTGVRTDEANTGTATPAPSPSCMRREAKTDDDD"
14960..15379
/ gene="OJ1341_A08.108"
<14960..15379
/ gene="OJ1341_A08.108"
/ note="start and end point are not identified"
14960..15379

```

```

/ gene="OJ1341_A08.108"
/ note="similar to Oryza sativa chromosome 4,
OSJNBa0050F15.5"
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83016.1"
/ db_xref="GI:34393703"
/ translation="MTKNLLBRADDDEHRLVALRRLTOAARRRDVPGNSDEMAVD
VAGASDALRAPSSTGPAGVLLLEQPADEFEEDGNDVDESETSVMSRGGCGDREGDNVQR
LGDITRRRGIEVAGEARTSQQVDVSPATDVSDERE"
complement (18618..19148)
/ gene="OJ1341_A08.109"
complement (<18618..19148)
/ gene="OJ1341_A08.109"

Query Match 85.0%; Score 17; DB 8; Length 67893;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
Db 33523 CGCGCCACCACTCGAC 33507

RESULT 16
AC135598 137312 bp DNA linear HTG 16-OCT-2003
LOCUS Oryza sativa (japonica cultivar-group) chromosome 3 clone
DEFINITION OSJNB0021K20, ** SEQUENCING IN PROGRESS **, 2 ordered pieces.
AC135598
VERSION AC135598.3 GI:37693584
KEYWORDS HTG; HTGS PHASE2; HTGS ACTIVEFIN.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 137312)
Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
Gansberger, K., Burgess, S., Jarrahi, B., Shvartsbeyn, M., Brenner, M.,
Ciecko, A., Pai, G., Varaken, S., Hansen, C., Utterbach, T.,
Feldlyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
Salzberg, S. and Fraser, C.
Oryza sativa ssp. japonica cv. Nipponbare OSJNB0021K20 BAC genomic
sequence
Unpublished
REFERENCE 2 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Oct 16, 2003 this sequence version replaced gi:24418097.
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 68692: contig of 68692 bp in length
* 68693 68792: gap of unknown length
* 68793 137312: contig of 68520 bp in length.
FEATURES
Location/Qualifiers
1..137312
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"

```


Db 75594 CGCCGACCACTCGAC 75610

RESULT 18
AE016783/c

LOCUS
AE016783.1

DEFINITION
Pseudomonas putida KT2440 section 10 of 21 of the complete genome.

ACCESSION
AE016783.1

VERSION
GI:26557027

KEYWORDS
Pseudomonas putida KT2440

SOURCE
Pseudomonas putida KT2440

ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 308015)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzes,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hohnsels,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoff,A., Tummel,B. and Fraser,C.
Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)

2 (bases 1 to 308015)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzes,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hohnsels,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoff,A., Tummel,B. and Fraser,C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

TITLE
Location/Qualifiers

JOURNAL
1. 308015
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
/note="similar to GP:5257182, and GP:5257182; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="esterified fatty acid cis/trans isomerase"
/protein_id="AAK67989.1"
/db_xref="GI:24983931"

AUTHORS
/translation="MWHRILAGAFALLISGAVFGQAPQSPALISYTRDIOPIETFEKVFACHNDAACQLKLESGEARGKAVPYQGRSXPVTRIFYDAHSEEQWRKKGFXSVLDNQGQALMARMLGHTPLTPNAKLPEEIVGLSRNMCPLEHFEADYAGAHPKEMPLAVGTGLTDKEYDMRWLAAGAPEVQIOPSAABARQAEELNRRGPTALVGRWLYEHLFLAHPAGGEGQHFQWVSRTPSGKPDVIAITRPNDDPGDGYRRLIPVQGVIVHKTHTITPMGQKLKRVQKLIYAGDWHNAALPGVGPFRHPRPPTFEALFAVARYQVFLMDNABYFVTRTFIRGVCRGQIATDVRNFWALFQEFADRIITDAKYRGAEATPLAMPQCIDVGSVLSLWHAYDKENDYEKLREAYAEVAPRWSTLWAGNDNALLSIFRHFDSVATKGLVGDVPLTWLFDYPTFERTYQLVNFDVYGNVSHQLQTRLYFDLIRNGAEVNFRLMPADQRKAILGDWQKSGKVMMDVEDIDTDPGSGIKLDPNPKRDFGLKLLQRTGSLNAAPDPINRCOGAFCSRCQMSSEFPNQAQSLRSLVSRPAAGKVINQLPENTMLRIGQDQGVQLSLNRAHSNVATLLGEAYVQPGCLDLTLTYPGVLSYKFNFIPTKQVDFEVEDMEYAKDDAAKFERIVNMGVRRKSHAPAFWRYHDLNSYKETEPEVEAGVLDNVRNENL"

FEATURES
source
1. 308015
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
/note="similar to GP:5257182, and GP:5257182; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="esterified fatty acid cis/trans isomerase"
/protein_id="AAK67989.1"
/db_xref="GI:24983931"

Gene
122..2422
/gene="cti"
/locus_tag="PP2376"

CDS
122..2422
/gene="cti"
/locus_tag="PP2376"

TITLE
Location/Qualifiers

JOURNAL
1. 308015
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
/note="similar to GP:5257182, and GP:5257182; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="esterified fatty acid cis/trans isomerase"
/protein_id="AAK67989.1"
/db_xref="GI:24983931"

AUTHORS
/translation="MWHRILAGAFALLISGAVFGQAPQSPALISYTRDIOPIETFEKVFACHNDAACQLKLESGEARGKAVPYQGRSXPVTRIFYDAHSEEQWRKKGFXSVLDNQGQALMARMLGHTPLTPNAKLPEEIVGLSRNMCPLEHFEADYAGAHPKEMPLAVGTGLTDKEYDMRWLAAGAPEVQIOPSAABARQAEELNRRGPTALVGRWLYEHLFLAHPAGGEGQHFQWVSRTPSGKPDVIAITRPNDDPGDGYRRLIPVQGVIVHKTHTITPMGQKLKRVQKLIYAGDWHNAALPGVGPFRHPRPPTFEALFAVARYQVFLMDNABYFVTRTFIRGVCRGQIATDVRNFWALFQEFADRIITDAKYRGAEATPLAMPQCIDVGSVLSLWHAYDKENDYEKLREAYAEVAPRWSTLWAGNDNALLSIFRHFDSVATKGLVGDVPLTWLFDYPTFERTYQLVNFDVYGNVSHQLQTRLYFDLIRNGAEVNFRLMPADQRKAILGDWQKSGKVMMDVEDIDTDPGSGIKLDPNPKRDFGLKLLQRTGSLNAAPDPINRCOGAFCSRCQMSSEFPNQAQSLRSLVSRPAAGKVINQLPENTMLRIGQDQGVQLSLNRAHSNVATLLGEAYVQPGCLDLTLTYPGVLSYKFNFIPTKQVDFEVEDMEYAKDDAAKFERIVNMGVRRKSHAPAFWRYHDLNSYKETEPEVEAGVLDNVRNENL"

Query Match
Best Local Similarity 100.0%; Score 17; DB 2; Length 137312;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
2 CGCCGACCACTCGAC 18
110026 CGCCGACCACTCGAC 110042

RESULT 17
CNS08C9S

LOCUS
CNS08C9S

DEFINITION
Oryza sativa chromosome 12. . BAC OSUNBa0040E18 of library OSUNBa from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza sativa (rice), complete sequence.

ACCESSION
AL772414

VERSION
AL772414.3

KEYWORDS
GI:24430251

SOURCE
Oryza sativa (japonica cultivar-group)

ORGANISM
Oryza sativa (japonica cultivar-group)

REFERENCE
1 (bases 1 to 156261)
Choi,S.N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P., Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M., Weissenbach,J. and Quetier,F.
Oryza sativa chromosome 12 sequencing
Unpublished
2 (bases 1 to 156261)
Direct Submission
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Oct 30, 2002 this sequence version replaced gi:23304718.
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

TITLE
Location/Qualifiers

JOURNAL
1. 158261
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OSUNBa0040E18"
/clone_lib="OSUNBa"

AUTHORS
/translation="MWHRILAGAFALLISGAVFGQAPQSPALISYTRDIOPIETFEKVFACHNDAACQLKLESGEARGKAVPYQGRSXPVTRIFYDAHSEEQWRKKGFXSVLDNQGQALMARMLGHTPLTPNAKLPEEIVGLSRNMCPLEHFEADYAGAHPKEMPLAVGTGLTDKEYDMRWLAAGAPEVQIOPSAABARQAEELNRRGPTALVGRWLYEHLFLAHPAGGEGQHFQWVSRTPSGKPDVIAITRPNDDPGDGYRRLIPVQGVIVHKTHTITPMGQKLKRVQKLIYAGDWHNAALPGVGPFRHPRPPTFEALFAVARYQVFLMDNABYFVTRTFIRGVCRGQIATDVRNFWALFQEFADRIITDAKYRGAEATPLAMPQCIDVGSVLSLWHAYDKENDYEKLREAYAEVAPRWSTLWAGNDNALLSIFRHFDSVATKGLVGDVPLTWLFDYPTFERTYQLVNFDVYGNVSHQLQTRLYFDLIRNGAEVNFRLMPADQRKAILGDWQKSGKVMMDVEDIDTDPGSGIKLDPNPKRDFGLKLLQRTGSLNAAPDPINRCOGAFCSRCQMSSEFPNQAQSLRSLVSRPAAGKVINQLPENTMLRIGQDQGVQLSLNRAHSNVATLLGEAYVQPGCLDLTLTYPGVLSYKFNFIPTKQVDFEVEDMEYAKDDAAKFERIVNMGVRRKSHAPAFWRYHDLNSYKETEPEVEAGVLDNVRNENL"

Query Match
Best Local Similarity 100.0%; Score 17; DB 8; Length 158261;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
2 CGCCGACCACTCGAC 18
110026 CGCCGACCACTCGAC 110042

```

CDS
2500...3651
/locus_tag="PP2377"
/translation="similar to GP:5257179, and GP:5257179; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acetyltransferase Act, putative"
/protein_id="AAN67990.1"
/db_xref="GI:24983932"
/translation="MQPAGPAPTGTGALLRLALASALDTGLNAPPSAGYSLYSL
QALLAFAMVVVCHHEMIDFEDFATGPIGOLLTDGAVGVDFIVISGLVILYSTRD
KTPRQFLRLALRIYPAVWFYALMAILLAPESOMPHOSFDRHLFLSLLFIPAE
NPGYGLPTLVNGWTINFEFFYLLPGLAFVLRQRHLLVYVALLVSEVLGKGV
VSRFYNNDDIIEFLLIGVGLVIRGLRQWLPALLGVAGNAIICLDSORLHLH
GLPSAMIVAFVMEPEFQGNRLIKALGDCSYVYLHVLYAGWAFASORLHLNPPFL
TFALCPVSGILMSWFSYQWLERGLYRRMQAWLAQGPAPAYALSRYNC"
3710...4294
/gene="yhg1"
/locus_tag="PP2378"
3710...4294
/gene="yhg1"
/locus_tag="PP2379"
/translation="similar to GB:M68840, GB:X60819, SP:P21397,
PID:187351, PID:187353, and PID:187355; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="yhg1 protein"
/protein_id="AAN67991.1"
/db_xref="GI:24983933"
/translation="MSAITIDAAHDYLDLLSKONTGIGIRIFITQPGTQVACTCI
AYCPGEKPDPTAVGLKSFAYLDVAVSPELEDAVDYATDRMGOLITKAPNAKVP
WYNDSPIRINYLQTEINPGLASHGQVSLVDVDDGIANVLQFGGCGCGCQADV
TLKGIERTLLERPELGVDRDVTDSQENAY"
4428...5099
/locus_tag="PP2379"
4428...5099
/locus_tag="PP2379"
/translation="similar to GP:3298363, SP:P04127, GB:X02921,
SP:P02972, SP:P04740, GB:M68060, and GB:X03391; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Scol/SenC family protein"
/protein_id="AAN67992.1"
/db_xref="GI:24983935"
/translation="NPRFDRPVRACCTFPVKDGFMDLLTRAVVAGMVLGILLIA
GCSFARGDEFKIGKMSNEILGRKFSKADPGQNVRTLSFFYGSFMVFFGTQCPAVC
PTTLARAQIRKLGRDRDLFQVVFITLDPEDTPEVLDAYKAFDPTFTALTGTPE
EIAAVAKFVFYKVPAGDTYTIHSSTSYVYDTRGLSLGHSLSNAKEADLVY
LMEIC"
5100...5582
/locus_tag="PP2380"
5100...5582
/locus_tag="PP2380"
/translation="similar to GB:D13178, SP:Q08751, PID:398125,
GB:D13178, SP:Q08751, and PID:398125; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN67993.1"
/db_xref="GI:24983936"
/translation="MSMQFIRKGLAALVLMGLALPALAQTTVSDAWRASVPHQOSTG
APMVLTAASDSKLIVGVSAVKTQVMGMVGMVGMREVKAIELPAGKAVITLDPNG
LHVLMGLHNQVKEDGKPLTLITIEDAKGETLEVQAVRALNADAGGGHMHNNH"
complement(5632...5874)
/locus_tag="PP2381"
complement(5632...5874)
/locus_tag="PP2381"
/codon_start=1
/transl_table=11

```

Query Match 85.0%; Score 17; DB 1; Length 308015;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCCGCACCCCTCGACT 19
|||||
Db 46664 GCCGCACCCCTCGACT 46648

RESULT 19
AF087589/c 276 bp DNA linear PLN 08-FEB-2000
LOCUS
DEFINITION Saxifraga marginata internal transcribed spacer 1, partial

```

/product="conserved domain protein"
/protein_id="AAN67994.1"
/db_xref="GI:24983937"
/translation="MEDPAMGERELTLLSLMQACQLSSACKIEADWIDROGDVPA
AGKIRASLKALEADEAQVRKLTLSLDRPLPRRS"
6229...7533
/locus_tag="PP2382"
6229...7533
/locus_tag="PP2382"
/translation="similar to GB:M11189, SP:P04190, SP:P14488, and
PID:142604; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transporter, putative"
/protein_id="AAN67995.1"
/db_xref="GI:24983938"
/translation="NAISIGSKCEPYESAIOVFRSDVARKYPLMLTLINFGTSLSL
GYSFKGCFEAFEAFAVAIVITDADMLLIGIAYFFRILPSVVFSLGWLADNANK
KHTLLAIEVTKGLLSPALFSPVFTVLLVGLAMVMTALECLYVPTFRAYFPDLV
EKDELATVNSGIQVIEDAASILGLPVFSACVLLSREATFLFFAMCLIVSAICITLT
PARCSVQAFDGCALIRDAAGSVSQLRAGNAPLEFVIGCTTLCAMFATSVIRFILPAS
VUEHFASAAVGVFVSELLAAGVVGCMLYTRNFETTAGLVIRYMLYVGLGFFFAAVA
LOFNWMLFILLFLVGF:GAFVDIAIVTINQCLSEHEVEGNFSLYIFTVMSORAVSG
LIASLVFVLGAPATFIWTLMLLIAPVRWNNKRGKRPDVSRT"
complement(7582...8427)
/locus_tag="PP2383"
complement(7582...8427)
/locus_tag="PP2383"
/translation="identified by match to PFAM protein family HMM
PF00165"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, AraC family"
/protein_id="AAN67996.1"
/db_xref="GI:24983939"
/translation="MSPPLREQTHLWQAPALCDVEMLHARYQORFAPRHVEGVYFTV
IESGAORFWHRGSEHLAPVGSVNLINDELHTGATAHEAGRWYRGFYDEHERVGVLD
EIELGRHGMPSFKDSVIHDPALAFSOLHQLSEASALOOQTAWROALVQRHG
QCAEPSPGHEPLELAVARELIESQLADPPSLEAALAAVNLSPHFARVFRQATGLPP
HAWLQORLARAREMLKHEGLAASQVAFDLGFADQSHLSQKQKQVGTGAYRQACVH
SALRA"
complement(8424...8738)
/locus_tag="PP2384"
complement(8424...8738)
/locus_tag="PP2384"
/translation="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN67997.1"
/db_xref="GI:24983940"
/translation="MTWILIFAMGAIVFLNRYAFLEPRLPURLSSNAQQLGFAVPGM
LTAICGPIIFMPDHQNLNLLNPYLLGSLVAIALVLTSLSLMLVSLMLIFFLLRSW
LA"
complement(8735...9427)
/gene="azlC"
/locus_tag="PP2385"
complement(8735...9427)
/gene="azlC"

```

```

sequence.
ACCESSION AF087589
VERSION AF087589.1 GI:5814034
SOURCE Saxifraga marginata
ORGANISM Saxifraga marginata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Saxifragales; Saxifragaceae; Saxifraga.
REFERENCE 1 (bases 1 to 276)
AUTHORS Conti, E., Soltis, D.E., Hardig, T.M. and Schneider, J.
TITLE Phylogenetic relationships of the silver saxifrages (Saxifraga,
sect. Ligulatae haworth): implications for the evolution of
substrate specificity, life histories, and biogeography
Mol. Phylogenet. Evol. 13 (3), 536-555 (1999)
JOURNAL 20088704
MEDLINE 10620412
PUBMED
REFERENCE 2 (bases 1 to 276)
AUTHORS Conti, E., Soltis, D.E., Michael, H. and Jason, S.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1998) Museum and Department of Biology and
Wildlife, University of Alaska-Fairbanks, 907 Yukon Drive,
Fairbanks, AK 99775-6960, USA
FEATURES
Source 1..276
Location/Qualifiers
misc_RNA /organism="Saxifraga marginata"
/db_type="genomic DNA"
/db_xref="taxon:102731"
<1..>276
/product="internal transcribed spacer 1"
ORIGIN
Query Match 84.0%; Score 16.8; DB 8; Length 276;
Best Local Similarity 90.0%; Pred. No. 4.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
|||||
Db 113 GCGCGCGCACCGCCGACTT 94

RESULT 20
CXSTPLUC1/c
LOCUS CXSTPLUC1 420 bp DNA linear BCT 22-AUG-1996
DEFINITION C.xyli DNA for strong promoter (420 bp).
ACCESSION X87993
VERSION X87993.1 GI:1321800
KEYWORDS hairpin loop; luciferase; repeated sequence; strong promoter.
SOURCE Leifsonia xyli
ORGANISM Leifsonia xyli
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococccineae; Microbacteriaceae; Leifsonia.
REFERENCE 1
AUTHORS Haapalainen, M., Karp, M. and Metzler, M.C.
TITLE Isolation of strong promoters from Clavibacter xyli subsp.
cynodontis using a promoter probe plasmid
Biochim. Biophys. Acta 1305 (3), 130-134 (1996)
JOURNAL 96180315
MEDLINE 8597597
PUBMED
REFERENCE 2 (bases 1 to 420)
AUTHORS Haapalainen, M.L.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-1995) M.L. Haapalainen, University of Turku,
Department of Biology, Biocity A 6, FIN-20520 Turku, FINLAND
FEATURES
Source 1..420
Location/Qualifiers
misc_structure 271..310
/organism="Leifsonia xyli"
/db_type="genomic DNA"
/strain="plasmidless, a natural isolate from Florida,
U.S."
/sub_species="cynodontis"
/db_xref="taxon:1575"

sequence.
repeat_unit /note="hairpin
putative"
275..282
/rpt_type=DIRECT
repeat_unit 317..324
/rpt_type=DIRECT
-35_signal 369..374
-10_signal 392..397
mRNA 401..>420
411..>420
ORIGIN
Query Match 84.0%; Score 16.8; DB 1; Length 420;
Best Local Similarity 90.0%; Pred. No. 4.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
|||||
Db 44 GCGCGCGCACCACTCGACTT 25

RESULT 21
AX122356/c
LOCUS AX122356 858 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 2272 from Patent EP1108790.
ACCESSION AX122356
VERSION AX122356.1 GI:14039309
KEYWORDS Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 2272 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source 1..858
/organism="Corynebacterium glutamicum"
/mol_type="unassigned DNA"
/db_xref="taxon:1718"
ORIGIN
Query Match 84.0%; Score 16.8; DB 6; Length 858;
Best Local Similarity 90.0%; Pred. No. 3.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
|||||
Db 36 GCGCGCGCACCACTCACTT 17

RESULT 22
BD164473/c
LOCUS BD164473 858 bp DNA linear PAT 17-JAN-2003
DEFINITION Novel polynucleotide.
ACCESSION BD164473
VERSION BD164473.1 GI:27870285
KEYWORDS JP 2002191370-A/2272.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 858)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2272 09-JUL-2002;
KYOWA HAKKO KOGYO CO LTD
COMMENT OS Corynebacterium glutamicum
PN JP 2002191370-A/2272
PD 09-JUL-2002

```


AF038597 1654 bp mRNA linear INV 22-DEC-1998
 LOCUS Drosophila melanogaster ornithine decarboxylase antizyme mRNA,
 complete cds.
 ACCESSION AF038597
 VERSION AF038597.1 GI:2708712
 KEYWORDS
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1654)
 AUTHORS Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F.,
 Gesteland, R.F. and Hayashi, S.
 TITLE Autoregulatory frameshifting in decoding mammalian ornithine
 decarboxylase antizyme
 JOURNAL Cell 80 (1), 51-60 (1995)
 MEDLINE 95112349
 PUBMED 7813017
 REFERENCE 2 (bases 1 to 1654)
 AUTHORS Ivanov, I.P., Simin, K., Letsou, A., Atkins, J.F. and Gesteland, R.F.
 TITLE The Drosophila gene for antizyme requires ribosomal frameshifting
 for expression and contains an intronic gene for snRNP Sm D3 on the
 opposite strand
 JOURNAL Mol. Cell. Biol. 18 (3) (1998) In press
 REFERENCE 3 (bases 1 to 1654)
 AUTHORS Ivanov, I.P., Simin, K., Letsou, A., Atkins, J.F. and Gesteland, R.F.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1997) Human Genetics, University of Utah, 15 N,
 2030 E, Room 6160, Salt Lake City, UT 84112-5330, USA
 FEATURES
 source
 location/Qualifiers
 1..1654
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /chromosome="2"
 join(147..329,331..912)
 /note="frameshifting which results in the products of ORF1
 and ORF2 being synthesized as a single polypeptide;
 analogous to that of the mammalian antizyme; the last two
 codons of ORF1 (tcc tga t) would be decoded as Ser-Asp
 through +1 frameshifting."
 /codon_start=1
 /product="ornithine decarboxylase antizyme"
 /protein_id="AAC97538.1"
 /db_xref="GI:2708713"
 /translation="MPSNNNNKLPVFSQGVRRFNDSDGIADGKLTITSTSSCATM
 SSSYRISLGVPLWMSDVPVHRTDHRASLTGYSRKSVDSAGGSLYEASRRSS
 LSSSQSPDLSPDTHSLCSDDCOEVLRLQHDOPVQITIKLHVTEOYTNNT
 ILAPVNNLLYALPKDLPAGSKQTFISLFEAEKLEVDGIWVMPKQPDARLIE
 AFLFMGFEPLSRAPQAPPAINDNENYFYLSIE"

Query Match 84.0%; Score 16.8; DB 3; Length 1654;
 Best Local Similarity 90.0%; Pred. No. 3.6e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GCGCGGCACACCTCGACTT 20
 |||||
 Db 239 GCTCCGCACCATCTCGACTT 258
 |||||

RESULT 25
 AY069220 1733 bp mRNA linear INV 17-DEC-2001
 LOCUS Drosophila melanogaster GH26763 full length cDNA.
 DEFINITION
 ACCESSION AY069220
 VERSION AY069220.1 GI:17861775
 KEYWORDS FLI CDNA.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

AF038597 1654 bp mRNA linear INV 22-DEC-1998
 LOCUS Drosophila melanogaster ornithine decarboxylase antizyme mRNA,
 complete cds.
 ACCESSION AF038597
 VERSION AF038597.1 GI:2708712
 KEYWORDS
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1654)
 AUTHORS Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F.,
 Gesteland, R.F. and Hayashi, S.
 TITLE Autoregulatory frameshifting in decoding mammalian ornithine
 decarboxylase antizyme
 JOURNAL Cell 80 (1), 51-60 (1995)
 MEDLINE 95112349
 PUBMED 7813017
 REFERENCE 2 (bases 1 to 1654)
 AUTHORS Ivanov, I.P., Simin, K., Letsou, A., Atkins, J.F. and Gesteland, R.F.
 TITLE The Drosophila gene for antizyme requires ribosomal frameshifting
 for expression and contains an intronic gene for snRNP Sm D3 on the
 opposite strand
 JOURNAL Mol. Cell. Biol. 18 (3) (1998) In press
 REFERENCE 3 (bases 1 to 1654)
 AUTHORS Ivanov, I.P., Simin, K., Letsou, A., Atkins, J.F. and Gesteland, R.F.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1997) Human Genetics, University of Utah, 15 N,
 2030 E, Room 6160, Salt Lake City, UT 84112-5330, USA
 FEATURES
 source
 location/Qualifiers
 1..1654
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /chromosome="2"
 join(147..329,331..912)
 /note="frameshifting which results in the products of ORF1
 and ORF2 being synthesized as a single polypeptide;
 analogous to that of the mammalian antizyme; the last two
 codons of ORF1 (tcc tga t) would be decoded as Ser-Asp
 through +1 frameshifting."
 /codon_start=1
 /product="ornithine decarboxylase antizyme"
 /protein_id="AAC97538.1"
 /db_xref="GI:2708713"
 /translation="MPSNNNNKLPVFSQGVRRFNDSDGIADGKLTITSTSSCATM
 SSSYRISLGVPLWMSDVPVHRTDHRASLTGYSRKSVDSAGGSLYEASRRSS
 LSSSQSPDLSPDTHSLCSDDCOEVLRLQHDOPVQITIKLHVTEOYTNNT
 ILAPVNNLLYALPKDLPAGSKQTFISLFEAEKLEVDGIWVMPKQPDARLIE
 AFLFMGFEPLSRAPQAPPAINDNENYFYLSIE"

Query Match 84.0%; Score 16.8; DB 3; Length 1654;
 Best Local Similarity 90.0%; Pred. No. 3.6e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GCGCGGCACACCTCGACTT 20
 |||||
 Db 239 GCTCCGCACCATCTCGACTT 258
 |||||

RESULT 25
 AY069220 1733 bp mRNA linear INV 17-DEC-2001
 LOCUS Drosophila melanogaster GH26763 full length cDNA.
 DEFINITION
 ACCESSION AY069220
 VERSION AY069220.1 GI:17861775
 KEYWORDS FLI CDNA.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1733)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,
 Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R.,
 Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C.J.,
 Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K.,
 Yu, C., Lewis, S.E., Rubin, G.M. and Celisner, S.
 TITLE Direct Submission
 JOURNAL Submitted (10-DEC-2001) Berkeley Drosophila Genome Project,
 Lawrence Berkeley National Laboratory, One Cyclotron Road,
 Berkeley, CA 94720, USA
 COMMENT Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory
 Berkeley, CA 94720
 This clone was sequenced as part of a high-throughput process to
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
 Science 2000). The sequence has been subjected to integrity checks
 for sequence accuracy, presence of a polyA tail and contiguity
 within 100 kb in the genome. Thus we believe the sequence to
 reflect accurately this particular cDNA clone. However, there are
 artifacts associated with the generation of cDNA clones that may
 have not been detected in our initial analyses such as internal
 priming, priming from contaminating genomic DNA, retained introns
 due to reverse transcription of unspliced precursor RNAs, and
 reverse transcriptase errors that result in single base changes.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our web site
 (http://fruitfly.berkeley.edu) or send email to
 cdna@fruitfly.berkeley.edu.
 location/Qualifiers
 1..1733
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /strain="Y; cn bw sp"
 /db_xref="taxon:7227"
 /map="48F1-48F1"
 1..1733
 /gene="guf"
 /note="alignment with genomic scaffold AE003823. Gene does
 not completely overlap longest ORF"
 /db_xref="FLYBASE:FBgn0014184"
 46..405
 /gene="guf"
 /note="Longest ORF"
 /codon_start=1
 /product="GH26763p"
 /protein_id="AAL39365.1"
 /db_xref="GI:17861776"
 /db_xref="FLYBASE:FBgn0014184"
 /translation="MHCGIAVHFRSASQNLGHALGQTOYIDRAHGVGLDGLHRIHV
 VNRWGRQCQVVDLDDLEKRIHYVVARAEFGTRRPEHSTTQIISEQSFHQWRKGT
 TTPPHTQISNRHFSM"

Query Match 84.0%; Score 16.8; DB 3; Length 1733;
 Best Local Similarity 90.0%; Pred. No. 3.6e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GCGCGGCACACCTCGACTT 20
 |||||
 Db 683 GCTCCGCACCATCTCGACTT 702
 |||||

RESULT 26
 AY071071 1755 bp mRNA linear INV 20-DEC-2001
 LOCUS Drosophila melanogaster RE14453 full length cDNA.
 DEFINITION
 ACCESSION AY071071
 VERSION AY071071.1 GI:17945271
 KEYWORDS FLI CDNA.
 SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1755)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M., and Celisner, S.

TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES
Source
1..1755
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="48F1-48F1"
gene
1..1755
/gene="guf"
/note="Alignment with genomic scaffold AE003823"
/db_xref="FLYBASE:FBgn0014184"
1358..1564
/gene="guf"
/note="Longest ORF"
/codon_start=1
/product="R3E14453p"
/protein_id="RAA48693.1"
/db_xref="GI:17945272"
/db_xref="FLYBASE:FBgn0014194"
/translation="MFSLINIFFMFPLYLCLSLYKVCNPLFALKNIQVIFH FLYABSDENGEVLEKDKGNAYA"

ORIGIN
Query Match 84.0%; Score 16.8; DB 3; Length 1755;
Best Local Similarity 90.0%; Pred. No. 3.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCCACCACTCGACTT 20
DB 244 GTCCGCACCACTCGACTT 263

RESULT 27
DMU29529
LOCUS Drosophila melanogaster gutfeeling (guf) mRNA, complete cds.
DEFINITION U29529
ACCESSION U29529
VERSION U29529.1 GI:1143044
KEYWORDS

ORGANISM Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 2163)
Salzberg, A., Golden, K., Bodmer, R. and Bellen, H.J.
gutfeeling, a Drosophila gene encoding an antizyme-like protein, is required for late differentiation of neurons and muscles
Genetics 144 (1), 183-196 (1996)
97032936
MEDLINE 8878684
PUBMED

REFERENCE 2 (bases 1 to 2163)
Salzberg, A., Golden, K., Bodmer, R. and Bellen, H.J.
AUTHORS Direct Submission
TITLE Submitted (19-JUN-1995) Adi Salzberg, HHMI, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA
JOURNAL
FEATURES
Location/Qualifiers
1..2163
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/chromosome="2"
/map="48E"
gene
1..2163
/gene="guf"
1261..1821
/gene="guf"
/standard_name="gutfeeling"
/note="Method: conceptual translation supplied by author; similar to human ornithine decarboxylase antizyme, encoded by Genbank Accession Number U09202"
/codon_start=1
/protein_id="AAB49330.1"
/db_xref="GI:1143045"
/translation="MLKCVNFRVGYSRKSSVDSAGSLYBASSRSLSSQSDCS LESOPDIHSLCSDDDCOEVLROILOHQPVOITIKLHVTEQYTNWNTILNPVNNLLY VALPKDLPPAGSKOTFIILFEAFKLEVDGIVMVPKQDPDRARLIEAFLFMGFPEL SRAQFAPPAINDNENTYFLYSIEE"

ORIGIN
Query Match 84.0%; Score 16.8; DB 3; Length 2163;
Best Local Similarity 90.0%; Pred. No. 3.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCCACCACTCGACTT 20
DB 1082 GTCCGCACCACTCGACTT 1101

RESULT 28
AE011952
LOCUS Xanthomonas axonopodis pv. citri str. 306
DEFINITION the complete genome.
ACCESSION AE011952 AE008923
VERSION AE011952.1 GI:21109405
KEYWORDS
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Xanthomonas axonopodis pv. citri str. 306
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.
REFERENCE 1 (bases 1 to 11366)
da Silva, A.C.R., Perro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chamergo, F., Cipina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Curcio-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Perro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Oksa, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and
 Kitajima, J.P.
 Comparison of the genomes of two Xanthomonas pathogens with
 differing host specificities
 Nature 417 (6887), 459-463 (2002)
 22022145
 12024217
 2 (bases 1 to 11366)
 da Silva, A.C.R., Ferro, J.A., Reinach, P.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
 Camargo, L.E.A., Canarotte, G., Cannavan, F., Cardoso, J.,
 Chambergo, F., Chapiana, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
 Furlino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Ferro, M.T., Formighieri, E.F., Franco, M.C.,
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A.,
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and
 Kitajima, J.P.
 Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
 São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,
 Brazil

FEATURES

Location/Qualifiers

1..11366
 /organism="Xanthomonas axonopodis pv. citri str. 306"
 /mol_type="genomic DNA"
 /strain="306"
 /db_xref="taxon:190486"
 /note="pathovar: citri"
 130..1971
 /gene="bga"
 /note="synonym: XAC3084"
 130..1971
 /gene="bga"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="beta-galactosidase"
 /protein_id="AA037929.1"
 /db_xref="GI:21109406"
 /translation="MLRTILALVLAFAIPITGTAETRWPNFTGQTQFVRDCK
 PYLLSGAIHFQIPRAYWKDLQKARALGLNTVTVFWNLVPEQOQDFDSGHNDV
 AAFVREAAGLNVLPQPCAEWAGGVPAWLGKGNIRVSRDPRFLAASQAVL
 DALANQVPLNNGGPIIAVOVENYGSYADDAHAYMADNRYVAGDFKALLFTSD
 GADMLANGTLPDUAUVNFAFGSAFGLKIFRPDPQPRVGEYWGAYWGWKGFYD
 AIDQAQAEFEFLRQHGSLANLYMFIGTSGFGMNGANFNQNSDHYAPQTTSYD
 AILDEAGPTPKALMDALARTVGVQPPALPAPIAITFTLFLRESLWNLNLPAP
 IAITDPQPGVGYLYLTVTTGPRKPLYLGDVRDVARVVDQRPVGSVERLL
 QVSLDVEIPAGQHTLVLVENSGRIYNGTRMADGRAGLVDPVLDNRLGTGWAPEL
 PMRPDSIRGWTRKAVQGPAPFHTGLIGTPTDYLDMRAFQKGFANGVNLGRHN
 IGPOTALYFAPFQKGNNTVVVFDLDDVAITPSVRLKQVWLKPKX"
 complement (2021..4945)
 /gene="XAC3085"
 complement (2021..4945)
 /gene="XAC3085"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037930.1"
 /db_xref="GI:21109407"
 /translation="MGVLLERGGSTAGLLCRRVLARHLRQLRAIHAAACILPIRRGQTQ

gene

CDS

gene

CDS

WRPAGINRTRWGAGAFVYGIACSSQQDWERTCAPSPCTGWSLLNVRVORSTQHL
 VTRDCAAVTPPAESQDQGWPLGALQTRPLRTHRDESSERVHHGDANARVA
 EETAYDVVAKYVADPLHGTADTAEILINDIPRRBPHEHSAEAGSAPPAKPAD
 HAAADPHSPHPQTAHPAQHAHAPHTPEGHQAGHAGHAPADQAVGHAKEGV
 LHALAEAGACAGADLMSGLTSGAMLPLSLIAVYAKREVAEQTHLRQERRRL
 REQALOSALCTAAPSAGAAHAAHALSEAIDTNAVLOARNADCAIATASMASAGVI
 PTKASELIGIOGLIAASKSANAAGLVGHSAAAGSAAAGIAGSVLAPLAPVAATA
 LGSTSLHQRERRARVAVDVGKRVFLQCLPESELSPGAQRYQYFVSTKLKIDRFAH
 RENQWNGFVVGITTYTATTKGVSAVAVAGATAAPVGTGLIIGAGLLGAATMGV
 GSHQFLLAGKOKRYRTQEDMGVDALLAVADLLPAREAESAAVOETPNVADQLQ
 PTALGVALTPTVAESAGSEEDIVAEADPGLEPKHROGSAAGAQAMPQPDAAAD
 ARAAGNAVPHGFELRSALYACIDGOKALDAFLQSGADDLQKLPKPRSTDSKGS
 GRAAHPLSRNVNAAALRAANYGRAVTVGKPHQARGKAMRSHAKHSAVLTETLAOW
 LQTPGSQKQIAYMOCCLQTKYLDKILGAQADLPQDVHDDPHSHGQTPVATVDRA
 PERLLAQHQRTRDQTAQVSLMLQELMAEQELANPDARQSKARAAWMPGLQRL
 IRLVTSNAIVQEGVAGFARFCMKQARQHTTHVTRGTLTATEVQAARIREQAAATPAS
 TT"
 5178..5378
 /gene="XAC3086"
 5178..5378
 /gene="XAC3086"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037931.1"
 /db_xref="GI:21109408"
 /translation="MSSSTVAQESVSAVAHTQNAFAEQIHQKPMGSAEAAHMTFLC
 QTESAKKPTLLPQMRDDDA"
 complement (5436..5765)
 /gene="rebB"
 /note="synonym: XAC3087"
 complement (5436..5765)
 /gene="rebB"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="rebB protein"
 /protein_id="AA037932.1"
 /db_xref="GI:21109409"
 /translation="VAFTAVNFOITDAVTQTNVVGGEAPAVAMSSIIQSMHSSGI
 LFQNAVAACQOQDITLAQAATNGVIQTYGVDTGACADVENVAQGAADNLNLSLTVL
 RSFDFAK"
 5784..6089
 /gene="XAC3088"
 5784..6089
 /gene="XAC3088"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA037933.1"
 /db_xref="GI:21109410"
 /translation="MDTRFVLCRPIGAKNASRLCSPHRRRTATPRSLCHHDSGRR
 RNSTUALFRFSARRNLGVIVRAGAVRLSHRCTEGFTAALMDHGRVHSRARVPGATS"
 6226..7386
 /gene="XAC3089"
 6226..7386
 /gene="XAC3089"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037934.1"
 /db_xref="GI:21109411"
 /translation="MHPRLSLRLVSLAASI.GAVVAQASPSPSPIMLSEFIADP
 APTQAHASTLLETDCGHLAAWFCGEHGRADVGINLAQRPQWQTPRRIADGAHA
 GVDGAVPAWNVLFQSATGVPOLYIKLGNPRQWNGRLSLTDCGASPPRQALQ
 ILGPIKNKPVQLNGRIILAPSSSDRGWRAHLESDDDGWQRQWQRLNDKALQAIQ
 PSLMYPGRLOALGRSQNKLFSTFLDGLHQQPMYLLLEVENSPSTDAVYLHGR
 ALLVYNPAIAGKMDWDRGRTLAVAVSNDGVHQRVLTLEDSAGDEFSPAVITQTRDGL

VHVSYTKRRRIKHVLDPTFRGECSPQAPVPTMLPERCTTACAHADPSS"
 7865..9358
 /gene="XAC3090"
 CDS 7865..9358
 /gene="XAC3090"
 /note="Identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="leucin rich protein"
 /protein_id="AA037935.1"
 /db_xref="GI:21109412"
 /translation="MPTVVGASPLRRSPAIRPYRDVLSQWRHYSADRNWHSAWROA
 NSNNPEVRRTRALKATDLELLEATAPARVALERSVPPQFPDQAFRUSHQHMFI
 DAAGMLPDTMQPAGLETTLARNPLRALPASIASLSRLRELSIACPELTELPH
 LASTDASGEHQGLVNLQSLRLTGITSLPASIANLQNLKSLKINSPLSALGPAIHQ
 LPKLELDLRLGCTALRSVPIFGSAPLKLKDCNLLTLPLDIHRLTQLEKLDLR
 GCVNLRLPSLIARLPANCIILVPHQAQDQHQVALEKVKALRIEDTQAQMLS
 TVIEERNPPLEGPSPYLPEKPPCTTTFGEIPLVQMLEESDPHELQVSDMAGP
 SPRIEESGRLHYTNVSNKQAQSAHLGIVDHLGCFVYHERASQDLAASLAKAVOM
 WKTELIYNAHPQDRTSPDTFTHIPEQVSDADDEQOTSPAPSRDQ"
 complement(9460..10257)
 /gene="cutC"
 /note="synonym: XAC3091"
 complement(9460..10257)
 /gene="cutC"
 /note="Identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="copper homeostasis protein"
 /protein_id="AA037935.1"
 /db_xref="GI:21109413"

Query Match 84.0%; Score 16.8; DB 1; Length 11366;
 Best Local Similarity 90.0%; Pred. No. 3.1e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCTCGACTT 20
 DB 9955 GGGCGGCACCTCGACTT 9974

RESULT 29
 AC015208 64195 bp DNA linear HTG 16-NOV-1999
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 DEFINITION AC015208
 ACCESSION AC015208.1 GI:6436127
 VERSION HTG; HTGS PHASE2
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 64195)
 Direct Submission
 Authors Adams, M. and Venter, J.C.
 Title Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 Comment This sequence was identified as CDM:10211252 by the submitter.
 For further information on this sequence e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..64195
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN
 1
 ..64195
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

Query Match 84.0%; Score 16.8; DB 2; Length 64195;
 Best Local Similarity 90.0%; Pred. No. 2.7e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCTCGACTT 20
 DB 41235 GCTCCGCACATCTCGACTT 41254

RESULT 30
 AP004394 98286 bp DNA linear HTG 21-MAR-2002
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 2 clone
 DEFINITION OJ1118_C06, *** SEQUENCING IN PROGRESS ***
 ACCESSION AP004394
 VERSION HTG; HTGS PHASE2
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.
 1
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
 clone-OJ1118_C06
 Published Only in Database (2001)
 2 (bases 1 to 98286)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Title Submitted (05-DEC-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7488
 The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..98286
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="OJ1118_C06"

ORIGIN
 1
 ..98286

Query Match 84.0%; Score 16.8; DB 2; Length 98286;
 Best Local Similarity 90.0%; Pred. No. 2.6e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCTCGACTT 20
 DB 33269 GGGCGGCACCTCGACTT 33250

Search completed: June 20, 2004, 11:42:53
 Job time : 564.069 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 114.921 Seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 gcgcgcaccacctcgactt 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	100.0	867	7 ACA38314	Aca38314 Prokaryot
C 2	20	100.0	870	7 ACA40752	Aca40752 Prokaryot
C 3	20	100.0	110000	4 AAI99682_32	Continuation (33 o
C 4	20	100.0	110000	4 AAI99682_32	Continuation (33 o
C 5	18.4	92.0	849	7 ACA39852	Aca39852 Prokaryot
C 6	18.4	92.0	849	7 ACA38167	Aca38167 Prokaryot
C 7	17	85.0	162	2 AAX18071	Aax18071 Coding se
C 8	17	85.0	176	2 AAX18072	Aax18072 Coding se
C 9	17	85.0	495	7 ACA44066	Aca44066 Prokaryot
C 10	16.8	84.0	858	5 AAH7237	Aah7237 C glutami
C 11	16.8	84.0	858	7 ACA00707	Aca00707 C. glutam
C 12	16.8	84.0	1089	6 ABQ91626	Abq91626 M. capsul
C 13	16.8	84.0	1154	4 ABL20151	Ab120151 Drosophil
C 14	16.8	84.0	1204	4 ABL18389	Ab118389 Drosophil
C 15	16.8	84.0	1207	4 ABL20153	Ab120153 Drosophil
C 16	16.8	84.0	1404	4 ABL20149	Ab120149 Drosophil
C 17	16.8	84.0	2310	4 ABL05611	Ab105611 Drosophil
C 18	16.8	84.0	3250	4 ABL4062	Ab14062 Drosophil
C 19	16.8	84.0	3489	4 ABL18389	Ab118389 Drosophil
C 20	16.8	84.0	5110	4 ABL06782	Ab106782 Drosophil
C 21	16.8	84.0	6190	4 ABL20148	Ab120148 Drosophil
C 22	16.8	84.0	7108	4 ABL05610	Ab105610 Drosophil
C 23	16.8	84.0	7393	4 ABL20150	Ab120150 Drosophil

C 24	16.8	84.0	9160	4 ABL20152	Ab120152 Drosophil
C 25	16.8	84.0	349980	5 AAH68531	Aah68531 C Glutami
C 26	16.4	82.0	3326	5 AAF74625	Aaf74625 Human GI
C 27	16.4	82.0	3587	7 ADA83747	Ada83747 Human GI
C 28	16.4	82.0	3600	4 AAD12302	Aad12302 Human CUB
C 29	16.4	82.0	3600	6 ABR30501	Ab30501 Human GI
C 30	16.4	82.0	3600	6 AAL45544	Aal45544 Human GI
C 31	16.4	82.0	3600	6 AAL45543	Aal45543 Human GI
C 32	16.4	82.0	3600	6 AAL45542	Aal45542 Human GI
C 33	16.4	82.0	3600	6 AAL45545	Aal45545 Human GI
C 34	16.4	82.0	3600	9 AAD19025	Ada19025 Human dis
C 35	16.4	82.0	8522	4 AAS27843	Aas27843 DNA encod
C 36	16.4	82.0	8522	9 ADB94646	Ad94646 Novel hum
C 37	16.4	82.0	14486	4 AAS27844	Aas27844 DNA encod
C 38	16.4	82.0	14486	9 ADB94647	Ad94647 Novel hum
C 39	16.2	81.0	677	3 AAC45084	Aac45084 Arabidops
C 40	16	80.0	293	6 ABL73207	Ab173207 Corn tase
C 41	16	80.0	416	3 AAC38335	Aac38335 Zea maye
C 42	16	80.0	1617	7 AAT23507	Aat23507 Cellobioh
C 43	15.8	79.0	446	8 ACH42306	Ach42306 Human toe
C 44	15.8	79.0	875	4 AAI59563	Aai59563 Human pol
C 45	15.8	79.0	894	7 ABE56529	Ab56529 C. violac
C 46	15.8	79.0	1117	4 AAS02402	Aas02402 Human sec
C 47	15.8	79.0	1449	4 ABL25167	Ab125167 Drosophil
C 48	15.8	79.0	1641	2 AAZ10823	Aaz10823 Choline o
C 49	15.8	79.0	2250	9 ADB63564	Ad63564 Human cDN
C 50	15.8	79.0	2706	7 ADA53073	Ada53073 Human cod
C 51	15.8	79.0	2837	6 ABR35667	Ab35667 cDNA sequ
C 52	15.8	79.0	3449	4 ABL25166	Ab125166 Drosophil
C 53	15.8	79.0	4069	4 AAH17772	Aah17772 Human cDN
C 54	15.8	79.0	5236	4 AAK15656	Aak15656 Human pol
C 55	15.8	79.0	5530	4 AAK52640	Aak52640 Human pol
C 56	15.8	79.0	8144	4 AAS46634	Aas46634 Tumour su
C 57	15.8	79.0	8913	7 ABZ10071	Abz10071 Haematopo
C 58	15.8	79.0	10519	6 ABE55009	Ab55009 Invertebr
C 59	15.8	79.0	11523	4 AAS45467	Aas45467 Chemical
C 60	15.8	79.0	11523	6 AAS46641	Aas46641 Tumour su
C 61	15.8	79.0	11523	6 ABL33922	Ab133922 Human imm
C 62	15.8	79.0	11523	6 ABR34001	Ab34001 Human DNA
C 63	15.8	79.0	11523	6 ABR28323	Ab28323 DNA trans
C 64	15.8	79.0	11523	7 ADA20382	Ada20382 Prostate
C 65	15.8	79.0	11523	7 ADB48189	Ad48189 Human ren
C 66	15.8	79.0	11913	9 ADB54093	Ab54093 Pretreat
C 67	15.4	77.0	114	3 AAA49078	Aa49078 Shut 2330
C 68	15.4	77.0	151	6 ABR25553	Ab25553 Human ORF
C 69	15.4	77.0	192	9 AUC87668	Auc87668 Human man
C 70	15.4	77.0	272	3 AAA49089	Aa49089 Piece 5 D
C 71	15.4	77.0	360	6 ABZ12519	Abz12519 Arabidops
C 72	15.4	77.0	504	6 ABQ38061	Abq38061 Oligonucl
C 73	15.4	77.0	504	6 ABQ38060	Abq38060 Oligonucl
C 74	15.4	77.0	547	3 AAC40683	Aac40683 Arabidops
C 75	15.4	77.0	596	3 AAC45054	Aac45054 Arabidops
C 76	15.4	77.0	598	3 AAC34034	Aac34034 Arabidops
C 77	15.4	77.0	625	6 ABQ42142	Abq42142 Oligonucl
C 78	15.4	77.0	625	6 ABQ42143	Abq42143 Oligonucl
C 79	15.4	77.0	633	4 ABL27051	Ab127051 Drosophil
C 80	15.4	77.0	861	7 ACA24986	Aca24986 Prokaryot
C 81	15.4	77.0	935	3 AAC46244	Aac46244 Arabidops
C 82	15.4	77.0	939	6 ABR54079	Ab54079 DNA encod
C 83	15.4	77.0	939	6 ABR54083	Ab54083 DNA encod
C 84	15.4	77.0	978	3 AAC39924	Aac39924 Arabidops
C 85	15.4	77.0	1038	3 AAA49097	Aa49097 synBX08-
C 86	15.4	77.0	1050	6 ABQ90321	Abq90321 M. capsul
C 87	15.4	77.0	1085	3 AAA96229	Aa96229 cDNA enco
C 88	15.4	77.0	1107	7 ABZ39173	Abz39173 N. gonorr
C 89	15.4	77.0	1344	7 ACA39628	Aca39628 Prokaryot
C 90	15.4	77.0	1401	4 AAS54457	Aas54457 Pseudomon
C 91	15.4	77.0	1401	7 ACA19506	Aca19506 Prokaryot
C 92	15.4	77.0	1588	7 ACA26952	Aca26952 Prokaryot
C 93	15.4	77.0	1779	5 AAS70967	Aas70967 DNA encod
C 94	15.4	77.0	2469	3 AAA49095	Aa49095 synBX08-
C 95	15.4	77.0	2652	4 AAH26468	Aah26468 Ancestral
C 96	15.4	77.0	2652	4 AAH26470	Aah26470 Semi-Opti

c 97	15.4	77.0	3172	4	ABL27050	Drosophil	15.2	76.0	30042	6	ABX09145	Abx09145 Mycobact		
c 98	15.4	77.0	4008	9	ADE07105	Novel cod	15.2	76.0	76363	9	ACF30938	Acf30938 Rice cult		
c 99	15.4	77.0	4014	7	ACD79241	Human AVA	172	76.0	86765	9	ADD14752	ADD14752 Human src		
c 100	15.4	77.0	4645	4	ABL14607	Drosophil	173	76.0	110000	2	AAV30458	AAV30458 Rhizobium		
c 101	15.4	77.0	5615	6	ABN5962	Novel hum	174	15.2	76.0	110000	2	AAV30459	AAV30459 Rhizobium	
c 102	15.4	77.0	5967	7	ACD79242	Novel AVA	175	15.2	76.0	110000	4	AAI99682	AAI99682-28	
c 103	15.4	77.0	6233	9	ADC87652	Human mam	176	15.2	76.0	110000	4	AAI99683	AAI99683-28	
c 104	15.4	77.0	7407	9	ADB78685	Human pot	177	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 105	15.4	77.0	7407	9	ADB78688	Human pot	178	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 106	15.4	77.0	7407	9	ADB78688	Human pot	179	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 107	15.4	77.0	7407	9	ADB78688	Human pot	180	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 108	15.4	77.0	7407	9	ADB78688	Human pot	181	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 109	15.4	77.0	7411	9	ADD29557	Human tum	182	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 110	15.4	77.0	7413	5	AA81747	N. mening	183	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 111	15.4	77.0	8059	3	AAA81747	Novel hum	184	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 112	15.4	77.0	8309	4	AAI59190	Novel hum	185	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 113	15.4	77.0	8309	8	ADB49173	Novel hum	186	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 114	15.4	77.0	20348	4	ABL14606	Drosophil	187	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 115	15.4	77.0	2185	3	AA63350	Streptomy	188	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 116	15.4	77.0	26970	9	ADC87695	Human mam	189	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 117	15.4	77.0	63164	3	AA63348	Streptomy	190	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 118	15.4	77.0	82993	6	ABX09140	Mycobacte	191	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 119	15.4	77.0	110000	3	AA81490	Novel hum	192	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 120	15.4	77.0	110000	3	AAI99682	Novel hum	193	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 121	15.4	77.0	110000	4	AAI99683	Novel hum	194	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 122	15.4	77.0	110000	4	AAI99683	Novel hum	195	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 123	15.4	77.0	110000	4	AAI99683	Novel hum	196	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 124	15.4	77.0	111393	7	AA54645	Streptomy	197	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 125	15.4	77.0	349980	3	AAI21608	Neisseria	198	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 126	15.4	77.0	76.0	309	6	ABQ98283	Human ORF	199	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool
c 127	15.2	76.0	354	7	ABX53473	Bovine ES	200	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 128	15.2	76.0	595	3	AA243029	Human 5'	201	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 129	15.2	76.0	654	3	AA27355	Prokaryot	202	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 130	15.2	76.0	735	6	AA24408	Human RNA	203	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 131	15.2	76.0	971	6	ABQ99417	Human cod	204	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 132	15.2	76.0	1086	7	AC42323	Prokaryot	205	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 133	15.2	76.0	1221	9	ADC36217	Weed cont	206	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 134	15.2	76.0	1347	7	ABX56047	M. echino	207	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 135	15.2	76.0	1374	9	ABE64800	Rat gene	208	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 136	15.2	76.0	1374	9	ABE63996	Rat gene	209	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 137	15.2	76.0	1391	3	AA02319	Human col	210	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 138	15.2	76.0	1451	9	ADC36218	Weed cont	211	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 139	15.2	76.0	1590	7	AC436218	Prokaryot	212	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 140	15.2	76.0	1843	4	ABL10155	Drosophil	213	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 141	15.2	76.0	1879	2	AAQ11126	Sequence	214	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 142	15.2	76.0	1959	7	AC437842	Prokaryot	215	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 143	15.2	76.0	2208	4	ABL23887	Drosophil	216	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 144	15.2	76.0	2244	9	ADB69883	C. neofo	217	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 145	15.2	76.0	2289	7	ACA51296	Prokaryot	218	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 146	15.2	76.0	2649	9	ADB69522	C. neofo	219	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 147	15.2	76.0	2694	7	ACA5743	Prokaryot	220	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 148	15.2	76.0	2887	3	AA56703	Human tra	221	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 149	15.2	76.0	2932	4	AD16103	Human int	222	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 150	15.2	76.0	2933	2	AAV59665	Human sec	223	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 151	15.2	76.0	2933	6	AB573652	Human cdn	224	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 152	15.2	76.0	2933	8	ACD82795	CDNA sequ	225	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 153	15.2	76.0	3195	7	AB233720	Human col	226	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 154	15.2	76.0	3276	6	AB573785	Human sec	227	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 155	15.2	76.0	3276	6	AB573785	Human sec	228	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 156	15.2	76.0	3276	8	ACD82928	CDNA sequ	229	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 157	15.2	76.0	3472	6	ABQ54759	Human ova	230	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 158	15.2	76.0	4307	4	ABL10154	Drosophil	231	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 159	15.2	76.0	4649	9	ADB69161	C. neofo	232	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 160	15.2	76.0	7812	3	AA261238	DNA encod	233	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 161	15.2	76.0	8391	7	ACA64735	Mycobacte	234	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 162	15.2	76.0	8391	7	ACA64735	Mycobacte	235	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 163	15.2	76.0	9098	5	AA544953	CDNA enco	236	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 164	15.2	76.0	9210	4	AAH52046	Mycobacte	237	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 165	15.2	76.0	9390	7	ACA45481	Prokaryot	238	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 166	15.2	76.0	16605	4	AAK87049	Human imm	239	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 167	15.2	76.0	16605	4	AAK87049	Human imm	240	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 168	15.2	76.0	18335	4	AA59578	Propionib	241	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	
c 169	15.2	76.0	18335	7	ACP64507	Propionib	242	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool	

C 243 14.8 74.0 2175 4 ABL97752 Human tes
 C 244 14.8 74.0 2247 7 ACA37588 Prokaryot
 C 245 14.8 74.0 2463 7 ACA25940 Prokaryot
 C 246 14.8 74.0 2466 7 ACA23983 Prokaryot
 C 247 14.8 74.0 2511 7 ACA39804 Prokaryot
 C 248 14.8 74.0 2761 4 ABL24718 Drosophil
 C 249 14.8 74.0 3083 7 ADA53528 Human cod
 C 250 14.8 74.0 3332 9 ADD44952 Rat gene
 C 251 14.8 74.0 3401 4 ABL20460 Drosophil
 C 252 14.8 74.0 3635 6 AAL45541 Murine Gl
 C 253 14.8 74.0 3668 9 ADD24893 DNA encod
 C 254 14.8 74.0 3789 2 AAT63514 Mycobacte
 C 255 14.8 74.0 3882 7 ABL37537 Streptomy
 C 256 14.8 74.0 4873 4 AAC91400 Brassica
 C 257 14.8 74.0 5110 5 AAS29201 Genomic s
 C 258 14.8 74.0 5110 6 ABA15205 Human ner
 C 259 14.8 74.0 5110 6 ABA15205 Human ner
 C 260 14.8 74.0 5110 6 ABA15205 Human ner
 C 261 14.8 74.0 5110 6 ABA15205 Human ner
 C 262 14.8 74.0 5110 6 ABA15205 Human ner
 C 263 14.8 74.0 5110 6 ABA15205 Human ner
 C 264 14.8 74.0 5110 6 ABA15205 Human ner
 C 265 14.8 74.0 5110 6 ABA15205 Human ner
 C 266 14.8 74.0 5110 6 ABA15205 Human ner
 C 267 14.8 74.0 5110 6 ABA15205 Human ner
 C 268 14.8 74.0 5110 6 ABA15205 Human ner
 C 269 14.8 74.0 5110 6 ABA15205 Human ner
 C 270 14.8 74.0 5110 6 ABA15205 Human ner
 C 271 14.8 74.0 5110 6 ABA15205 Human ner
 C 272 14.8 74.0 5110 6 ABA15205 Human ner
 C 273 14.8 74.0 5110 6 ABA15205 Human ner
 C 274 14.8 74.0 5110 6 ABA15205 Human ner
 C 275 14.8 74.0 5110 6 ABA15205 Human ner
 C 276 14.8 74.0 5110 6 ABA15205 Human ner
 C 277 14.8 74.0 5110 6 ABA15205 Human ner
 C 278 14.8 74.0 5110 6 ABA15205 Human ner
 C 279 14.8 74.0 5110 6 ABA15205 Human ner
 C 280 14.8 74.0 5110 6 ABA15205 Human ner
 C 281 14.8 74.0 5110 6 ABA15205 Human ner
 C 282 14.8 74.0 5110 6 ABA15205 Human ner
 C 283 14.8 74.0 5110 6 ABA15205 Human ner
 C 284 14.8 74.0 5110 6 ABA15205 Human ner
 C 285 14.8 74.0 5110 6 ABA15205 Human ner
 C 286 14.8 74.0 5110 6 ABA15205 Human ner
 C 287 14.8 74.0 5110 6 ABA15205 Human ner
 C 288 14.8 74.0 5110 6 ABA15205 Human ner
 C 289 14.8 74.0 5110 6 ABA15205 Human ner
 C 290 14.8 74.0 5110 6 ABA15205 Human ner
 C 291 14.8 74.0 5110 6 ABA15205 Human ner
 C 292 14.8 74.0 5110 6 ABA15205 Human ner
 C 293 14.8 74.0 5110 6 ABA15205 Human ner
 C 294 14.8 74.0 5110 6 ABA15205 Human ner
 C 295 14.8 74.0 5110 6 ABA15205 Human ner
 C 296 14.8 74.0 5110 6 ABA15205 Human ner
 C 297 14.8 74.0 5110 6 ABA15205 Human ner
 C 298 14.8 74.0 5110 6 ABA15205 Human ner
 C 299 14.8 74.0 5110 6 ABA15205 Human ner
 C 300 14.8 74.0 5110 6 ABA15205 Human ner

ALIGNMENTS

RESULT 1
 ID ACA38314/C
 AC ACA38314;
 DT 19-JUN-2003 (first entry)
 XX Prokaryotic essential gene #19971.

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 drug design; gene.
 XX Mycobacterium bovis.
 OS WO200277183-A2.
 PN 03-OCT-2002.
 PD 21-MAR-2002; 2002WO-US009107.
 PF 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923F.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR WPI: 2003-029926/02.
 DR P-PSDB; ABU34444.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 14; SEQ ID NO 26184; 1766pp; English.
 The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway;
 (8) required for proliferation, or that inhibits cellular proliferation; (9)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is one of the target
 prokaryotic essential genes. Note: The sequence data for this patent did
 not form part of the printed specification, but was obtained in
 electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

Query Match 100.0%; Score 20; DB 7; Length 867;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCCGACACCTCGACTT 20
 DB 36 GCGCCGACACCTCGACTT 17

RESULT 2

ACA40752/c

ID ACA40752 standard; DNA; 870 BP.

XX AC

XX ACA40752;

XX DT

XX 19-JUN-2003 (first entry)

XX DE

XX Prokaryotic essential gene #22409.

XX KW

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KM

XX drug design; gene.

XX OS

XX Mycobacterium tuberculosis.

XX PN

XX WO200277183-A2.

XX PD

XX 03-OCT-2002.

XX PF

XX 21-MAR-2002; 2002WO-US009107.

XX PR

XX 21-MAR-2001; 2001US-00815242.

XX PR

XX 06-SEP-2001; 2001US-00948993.

XX PR

XX 25-OCT-2001; 2001US-0342923P.

XX PR

XX 08-FEB-2002; 2002US-00072851.

XX PR

XX 06-MAR-2002; 2002US-0362699P.

XX PA

XX (ELIT-) ELITRA PHARM INC.

XX XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI

XX Wall L, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR

XX WPI: 2003-029926/02.

XX DR

XX P-PSDB; ABU36882.

XX XX

XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT

XX for homologous nucleic acids required for cellular proliferation to

XX PT

XX isolate candidate molecules for rational drug discovery programs.

XX XX

XX Claim 14; SEQ ID NO 28622; 1766pp; English.

XX XX

XX The invention relates to an isolated nucleic acid comprising any one of

XX CC

XX the 6213 antisense sequences given in the specification where expression

XX CC

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC

XX encoding a polypeptide whose expression is inhibited by the antisense

XX CC

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC

XX polypeptide or its fragment whose expression is inhibited by the

XX CC

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC

XX proliferation or the activity of a gene in an operon required for

XX CC

XX proliferation; (7) identifying a compound that influences the activity of

XX CC

XX the gene product or that has an activity against a biological pathway

XX CC

XX required for proliferation, or that inhibits cellular proliferation; (8)

XX CC

XX identifying a gene required for cellular proliferation or the biological

XX CC

XX pathway in which a proliferation-required gene or its gene product lies

XX CC

XX on a gene on which the test compound that inhibits proliferation of an

XX CC

XX organism's activity; (9) manufacturing an antibiotic; (10) profiling a

XX CC

XX product is overexpressed or underexpressed; (12) determining the extent

XX CC

XX to which each of the strains is present in a culture or collection of

XX CC

XX strains; or (13) identifying the target of a compound that inhibits the

XX CC

XX proliferation of an organism. The antisense nucleic acids are useful for

XX CC

XX identifying proteins or screening for homologous nucleic acids required

XX CC

XX for cellular proliferation to isolate candidate molecules for rational

XX CC

XX drug discovery programs, or for screening homologous nucleic acids

XX CC

XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC

XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

XX CC

XX prokaryotic essential genes. Note: the sequence data for this patent did

XX CC

XX not form part of the printed specification, but was obtained in

XX CC

XX electronic format directly from WIPO at

XX CC

XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 870 BP; 137 A; 269 C; 315 G; 149 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 7; Length 870;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20

Db 36 GGGCGGCACCACTCGACTT 17

RESULT 3

AAI99682_12

Continuation (33 of 45) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain

WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP Fragment Name Begin End

WP AAI99682_00 1 110000

WP AAI99682_01 100001 210000

WP AAI99682_02 200001 310000

WP AAI99682_03 300001 410000

WP AAI99682_04 400001 510000

WP AAI99682_05 500001 610000

WP AAI99682_06 600001 710000

WP AAI99682_07 700001 810000

WP AAI99682_08 800001 910000

WP AAI99682_09 900001 1010000

WP AAI99682_10 1000001 1110000

WP AAI99682_11 1100001 1210000

WP AAI99682_12 1200001 1310000

WP AAI99682_13 1300001 1410000

WP AAI99682_14 1400001 1510000

WP AAI99682_15 1500001 1610000

WP AAI99682_16 1600001 1710000

WP AAI99682_17 1700001 1810000

WP AAI99682_18 1800001 1910000

WP AAI99682_19 1900001 2010000

WP AAI99682_20 2000001 2110000

WP AAI99682_21 2100001 2210000

WP AAI99682_22 2200001 2310000

WP AAI99682_23 2300001 2410000

WP AAI99682_24 2400001 2510000

WP AAI99682_25 2500001 2610000

WP AAI99682_26 2600001 2710000

WP AAI99682_27 2700001 2810000

WP AAI99682_28 2800001 2910000

WP AAI99682_29 2900001 3010000

WP AAI99682_30 3000001 3110000

WP AAI99682_31 3100001 3210000

WP AAI99682_32 3200001 3310000

WP AAI99682_33 3300001 3410000

WP AAI99682_34 3400001 3510000

WP AAI99682_35 3500001 3610000

WP AAI99682_36 3600001 3710000

WP AAI99682_37 3700001 3810000

WP AAI99682_38 3800001 3910000

WP AAI99682_39 3900001 4010000

WP AAI99682_40 4000001 4110000

WP AAI99682_41 4100001 4210000

WP AAI99682_42 4200001 4310000

WP AAI99682_43 4300001 4410000

WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 20; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20

Db 39432 GGGCGGCACCACTCGACTT 39451

RESULT 4

AAI199683 32
Continuation (33 of 44) of AAI199683 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI199683 Accession AAI199683
WP Fragment Name Begin End
WP AAI199683_00 1 110000
WP AAI199683_01 100001 210000
WP AAI199683_02 200001 310000
WP AAI199683_03 300001 410000
WP AAI199683_04 400001 510000
WP AAI199683_05 500001 610000
WP AAI199683_06 600001 710000
WP AAI199683_07 700001 810000
WP AAI199683_08 800001 910000
WP AAI199683_09 900001 1010000
WP AAI199683_10 1000001 1110000
WP AAI199683_11 1100001 1210000
WP AAI199683_12 1200001 1310000
WP AAI199683_13 1300001 1410000
WP AAI199683_14 1400001 1510000
WP AAI199683_15 1500001 1610000
WP AAI199683_16 1600001 1710000
WP AAI199683_17 1700001 1810000
WP AAI199683_18 1800001 1910000
WP AAI199683_19 1900001 2010000
WP AAI199683_20 2000001 2110000
WP AAI199683_21 2100001 2210000
WP AAI199683_22 2200001 2310000
WP AAI199683_23 2300001 2410000
WP AAI199683_24 2400001 2510000
WP AAI199683_25 2500001 2610000
WP AAI199683_26 2600001 2710000
WP AAI199683_27 2700001 2810000
WP AAI199683_28 2800001 2910000
WP AAI199683_29 2900001 3010000
WP AAI199683_30 3000001 3110000
WP AAI199683_31 3100001 3210000
WP AAI199683_32 3200001 3310000
WP AAI199683_33 3300001 3410000
WP AAI199683_34 3400001 3510000
WP AAI199683_35 3500001 3610000
WP AAI199683_36 3600001 3710000
WP AAI199683_37 3700001 3810000
WP AAI199683_38 3800001 3910000
WP AAI199683_39 3900001 4010000
WP AAI199683_40 4000001 4110000
WP AAI199683_41 4100001 4210000
WP AAI199683_42 4200001 4310000
WP AAI199683_43 4300001 4403765

Query Match 100.0%; Score 20; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGCGCACCACCTCGACTT 20
|||||
DB 33708 GCGCGCACCACCTCGACTT 33727

RESULT 5
ACA39852/c
ID ACA39852 standard; DNA; 849 BP.
XX AC
XX ACA39852;
XX
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #21509.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Mycobacterium leprae.
XX OS

PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR F-PSDB; ABU35982.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 27722; 17665p; English.
PS
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC required for proliferation, or that inhibits cellular proliferation; (9)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 849 BP; 150 A; 236 C; 279 G; 184 T; 0 U; 0 Other;

Query Match 92.0%; Score 18.4; DB 7; Length 849;
Best Local Similarity 95.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGCGCACCACCTCGACTT 20
|||||
DB 36 GCGCGCACCACCTCGACTT 17

RESULT 6
ACA38167/c
ID ACA38167 standard; DNA; 849 BP.
XX

AC ACA38167;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #19824.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Mycobacterium avium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR P-PSDB; ABU34297.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 26037; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 849 BP; 128 A; 301 C; 297 G; 121 T; 0 U; 2 Other;
Query Match 92.0%; Score 18.4; DB 7; Length 849;
Best Local Similarity 95.0%; Pred. NO. 98;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCTGACTT 20
DB 36 GCGCGCGCACCTGACTT 17
RESULT 7
AA18071/c
ID AAX18071 standard; DNA; 162 BP.
XX
AC AAX18071;
XX
DT 04-MAY-1999 (first entry)
XX
DE Coding sequence for human D2H binding protein DAB3.
XX
KW Gastro-intestinal transport receptor; binding protein; hsr; HPT1; D2H;
KW hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US010089.
XX
PR 15-MAY-1997; 97US-0046595P.
XX
XX (CYTO-) CYTOGEN CORP.
XX (ELAN-) ELAN CORP PLC.
XX
XX Alvarez VL, Omahony DJ, Lambkin IJ, Patterson CA, Singleton J;
PI Belinka BA, Carter JM, Cagney GW;
XX
XX WPI; 1999-009568/01.
XX
PT New proteins that bind specifically to receptors in the gastro-intestinal
PT tract and related nucleic acid-chimaeras and antibodies, used to
PT deliver therapeutic or diagnostic agents to, or through, the
PT gastrointestinal tract, e.g. insulin or leuprolide.
XX
PS Claim 49; Page 57; 294pp; English.
XX
XX This sequence encodes a peptide that specifically binds to the human D2H
CC protein. The invention relates to purified proteins (I) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1), hPEPT1,
CC D2H and human sucrose-isomaltase complex (hsr). (I) provide active
CC transport of therapeutic agents through human and animal GI tissue (into
CC the blood) for in vivo delivery, particularly for treatment or prevention
CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation
XX
SQ Sequence 162 BP; 30 A; 58 C; 42 G; 32 T; 0 U; 0 Other;
Query Match 85.0%; Score 17; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 495 BP; 67 A; 145 C; 168 G; 115 T; 0 U; 0 Other;

Query Match 85.0%; Score 17; DB 7; Length 495;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCGCGCACCTCGACT 19
DB 277 GCGCGCACCTCGACT 261

RESULT 10
AAH67237/c
ID AAH67237 standard; DNA; 858 BP.

XX AC AAH67237;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 2272.

XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-00127698.

XX PR 16-DEC-1999; 99JP-00377484.

XX PR 07-APR-2000; 2000JP-00159162.

XX PR 03-AUG-2000; 2000JP-00280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR P-PSDB; AAG92018.

XX PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analyzing

PT expression profile or pattern of a gene and identifying homologous gene.

PS Claim 8; SEQ ID NO 2272; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium, measuring expression amount and analysing

CC the expression profile or expression pattern of a gene derived from

CC Corynebacterium, and identifying a homologue of a gene derived from

CC corynebacterium. Corynebacterium bacteria are useful for producing amino

CC acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention. Note: The sequence data for this

CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 495 BP; 67 A; 145 C; 168 G; 115 T; 0 U; 0 Other;

Query Match 85.0%; Score 17; DB 7; Length 495;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCGCGCACCTCGACT 19
DB 277 GCGCGCACCTCGACT 261

RESULT 10
AAH67237/c
ID AAH67237 standard; DNA; 858 BP.

XX AC AAH67237;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 2272.

XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-00127698.

XX PR 16-DEC-1999; 99JP-00377484.

XX PR 07-APR-2000; 2000JP-00159162.

XX PR 03-AUG-2000; 2000JP-00280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR P-PSDB; AAG92018.

XX PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analyzing

PT expression profile or pattern of a gene and identifying homologous gene.

PS Claim 8; SEQ ID NO 2272; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium, measuring expression amount and analysing

CC the expression profile or expression pattern of a gene derived from

CC Corynebacterium, and identifying a homologue of a gene derived from

CC corynebacterium. Corynebacterium bacteria are useful for producing amino

CC acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention. Note: The sequence data for this

DB 36 GCGCGCACCTCGACTT 17

RESULT 11
ACAO0707/c
ID ACA00707 standard; DNA; 858 BP.
XX AC ACA00707;

XX DT 03-JUN-2003 (first entry)

XX DE C. glutamicum derived ORF SEQ ID 698.

XX KW Coryneform; nucleic acid array; fermentation; culture; ds.

XX OS Corynebacterium glutamicum.

XX PN DE10128510-A1.

XX PD 19-DEC-2002.

XX PF 13-JUN-2001; 2001DE-01028510.

XX PR 13-JUN-2001; 2001DE-01028510.

XX PA (DEGS) DEGUSSA AG.

XX PI Farwick M, Moeckel B, Pfefferle W, Bathe B, Huthmacher K;

XX PD WPI; 2003-279970/28.

XX PT New nucleic acid array useful for monitoring mRNA expression of
PT Corynebacterium glutamicum during fermentation, comprising nucleic acid
PT from Corynebacterium glutamicum.

XX PS Claim 1; Page 285; 709pp; German.

XX CC This invention describes a novel nucleic acid array involving
CC Corynebacterium glutamicum polynucleotides. The arrays are used to
CC analyse C. glutamicum, particularly for monitoring a fermentation process
CC to determine expression levels of C. glutamicum cellular mRNA. Such
CC monitoring particularly differentiates between expression levels of
CC different strains of C. glutamicum and allows the adjustment of different
CC culture and fermentation conditions. ACA00010-ACA02188 represent C.
CC glutamicum derived polynucleotides described in the disclosure of the
CC invention

XX SQ Sequence 858 BP; 202 A; 242 C; 225 G; 189 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 7; Length 858;

Best Local Similarity 90.0%; Pred. No. 4.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACTT 20

DB 36 GCGCGCACCTCGACTT 17

RESULT 12

ABQ91626

ID ABQ91626 standard; DNA; 1089 BP.

XX AC ABQ91626;

XX DT 01-OCT-2002 (first entry)

XX DE M. capsulatus gene #1611 for DNA array.

XX KW Micro array; gene; ds; differential expression; gene expression.

XX OS Methylococcus capsulatus.

XX

QY 1 GCGCGCACCTCGACTT 20

PN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002WO-N0000019.
XX
PR 12-JAN-2001; 2001NO-00000235.
PR 12-JAN-2001; 2001NO-00000239.
XX
PA (UNIF-) UNIFOB STIFTTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Bisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557818/59.
XX
PT Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes.
XX
PS Claim 14; Page 613; 678pp; English.
XX
CC The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC invention
XX
SQ Sequence 1089 BP; 260 A; 321 C; 338 G; 170 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 6; Length 1089;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 870 GCGCGCACCACTCGACTT 889
RESULT 13
ABL20151
ID ABL20151 standard; DNA; 1154 BP.
XX
AC ABL20151;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11926.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 11926; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1154 BP; 285 A; 310 C; 277 G; 282 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 4; Length 1154;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 159 GCTCCGCACCACTCGACTT 178
RESULT 14
ABL18389
ID ABL18389 standard; DNA; 1204 BP.
XX
AC ABL18389;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6640.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 6640; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 11926; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1154 BP; 285 A; 310 C; 277 G; 282 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 4; Length 1154;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 159 GCTCCGCACCACTCGACTT 178
RESULT 14
ABL18389
ID ABL18389 standard; DNA; 1204 BP.
XX
AC ABL18389;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6640.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 6640; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1204 BP; 286 A; 321 C; 288 G; 309 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 1204;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
|||
DB 209 GCTCCGCACCATCTCGACTT 228

RESULT 15
ABL20153
ID ABL20153 standard; DNA; 1207 BP.

XX ABL20153;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11932.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.

XX Claim 1; SEQ ID NO 11932; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1207 BP; 304 A; 320 C; 283 G; 300 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 1207;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
|||
DB 212 GCTCCGCACCATCTCGACTT 231

RESULT 16

ABL20149

ID ABL20149 standard; DNA; 1404 BP.

XX ABL20149;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11920.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.

XX Claim 1; SEQ ID NO 11920; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1404 BP; 373 A; 381 C; 339 G; 311 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 1404;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
|||
DB 409 GCTCCGCACCATCTCGACTT 428

RESULT 17

ABL05611/c

ID ABL05611 standard; cDNA; 2310 BP.

XX ABL05611;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11315.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

```

XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB61508.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 11315; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2310 BP; 608 A; 754 C; 571 G; 377 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 4; Length 2310;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 GCGCGCGCACCTCGACTT 20
Db 1920 GCGCGCGCACCTCGACTT 1901
RESULT 18
ABL14062/c
ID ABL14062 standard; cDNA; 3250 BP.
XX AC ABL14062;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36668.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX FN WO200171042-A2.
XX XX 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 11315; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2310 BP; 608 A; 754 C; 571 G; 377 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 4; Length 2310;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 GCGCGCGCACCTCGACTT 20
Db 1920 GCGCGCGCACCTCGACTT 1901

```

```

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 36668; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 3250 BP; 975 A; 697 C; 755 G; 823 T; 0 U; 0 Other;
Query Match 84.0%; Score 16.8; DB 4; Length 3250;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 GCGCGCGCACCTCGACTT 20
Db 378 GCTCCGCGCACCTCGACTT 359
RESULT 19
ABL18388/c
ID ABL18388 standard; DNA; 3489 BP.
XX AC ABL18388;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6637.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX FN WO200171042-A2.
XX XX 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 6637; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABB72072). The sequence data for this patent did not form part of the

```

CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 3489 BP; 1067 A; 715 C; 723 G; 984 T; 0 U; 0 Other;
 Query Match 84.0%; Score 16.8; DB 4; Length 3489;
 Best Local Similarity 90.0%; Pred. No. 4.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
 DB 2062 GCTCCGACCACTCGACTT 2043

RESULT 20
 ABL06782/c
 ID ABL06782 standard; cDNA; 5110 BP.

XX ABL06782;
 AC ABL06782;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14828.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB62679.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 14828; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 5110 BP; 1321 A; 1226 C; 1260 G; 1303 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 5110;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20

DB 302 GCGCCGACCACTCGACTT 283

RESULT 21
 ABL20148/c
 ID ABL20148 standard; DNA; 6190 BP.

XX ABL20148;

AC ABL20148;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11917.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 11917; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 6190 BP; 1866 A; 1245 C; 1340 G; 1739 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 6190;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20

DB 2062 GCTCCGACCACTCGACTT 2043

RESULT 22
 ABL05610/c

ID ABL05610 standard; cDNA; 7108 BP.

XX ABL05610;

AC ABL05610;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11312.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR P-FSDB; ABB61507.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11312; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7108 BP; 1909 A; 1824 C; 1701 G; 1674 T; 0 U; 0 Other;
 SQ
 Query Match 84.0%; Score 16.8; DB 4; Length 7108;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCGCACCACTCGACTT 20
 DB 5718 GCGCGCGCACCACTCGACTT 5699
 RESULT 23
 ABL20150/c
 ID ABL20150 standard; DNA; 7393 BP.
 XX
 AC ABL20150;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11923.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11929; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7108 BP; 1909 A; 1824 C; 1701 G; 1674 T; 0 U; 0 Other;
 SQ

XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11923; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7393 BP; 2294 A; 1463 C; 1583 G; 2053 T; 0 U; 0 Other;
 SQ
 Query Match 84.0%; Score 16.8; DB 4; Length 7393;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCGCACCACTCGACTT 20
 DB 2062 GCTCGCGCACCACTCGACTT 2043
 RESULT 24
 ABL20152/c
 ID ABL20152 standard; DNA; 9160 BP.
 XX
 AC ABL20152;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11929.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11929; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the

XX DE Human GLI gene.
XX KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;
XX KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
XX KW vaccine; ds; gene.
XX OS Homo sapiens.
XX XX
XX PN WO2002103028-A2.
XX PD 27-DEC-2002.
XX XX
XX PF 30-MAY-2002; 2002WO-IB004189.
XX XX
XX PR 30-MAY-2001; 2001US-0293999P.
XX PR 22-OCT-2001; 2001US-0330457P.
XX PR 19-FEB-2002; 2002US-0357144P.
XX XX
XX PA (BIOM-) BIOMEDICAL CENT.
XX XX
XX PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;
XX DR WPI: 2003-175241/17.
XX DR P-PSDB; ADA83748.
XX XX
XX PT Determining if a nucleic acid is a marker for a phenotype/cell type of
XX PT interest, by global comparison of expressed sequence tags known to be
XX PT expressed in the phenotype/cell type with all ESTs expressed in normal
XX PT tissue.
XX PS Claim 23; Page 104-106; 516pp; English.
XX CC The invention relates to a novel method for determining if a nucleic acid
XX CC is a marker for a predetermined phenotype/cell type of interest from a
XX CC biological species. The method comprises performing a global comparison
XX CC of a group of expressed sequence tags (ESTs) known to be expressed in the
XX CC phenotype/cell type of interest with all ESTs expressed in normal tissue
XX CC in order to identify ESTs that are preferentially expressed in the
XX CC phenotype/cell of interest. A method of the invention is useful for
XX CC determining whether a nucleic acid is a marker for a predetermined
XX CC phenotype or cell type of interest from a biological species, preferably
XX CC Arabidopsis or human. The cell type of interest is an abnormal cell such
XX CC as a tumour cell, and the predetermined phenotype is a stress-induced
XX CC phenotype such as hyperosmotic stress or high salt conditions. A method
XX CC of the invention is also useful for determining the progression of colon
XX CC cancer in a human, for detecting a tumour cell, and for regulating or
XX CC preventing the growth of a tumour cell. An antibody of the invention is
XX CC useful for detecting the absence or presence of peptides encoded by
XX CC tumour-associated markers. A polypeptide of the invention is useful as an
XX CC immunogen for vaccinating an animal. The present sequence encodes a
XX CC tumour-associated antigen of the invention.
XX SQ Sequence 3587 BP; 798 A; 1143 C; 944 G; 702 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 7; Length 3587;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCGCACACCTCGACTT 20
DB 3003 GCAGCACCACCTCGACTT 3020

RESULT 28
RAD12302
ID AAD12302 standard; cDNA; 3600 BP.
XX AC AAD12302;
XX XX
XX DT 16-OCT-2001 (first entry)
XX DE Human Cubitus interruptus (Ci) homologue, GLI-1 cDNA.

Query Match 82.0%; Score 16.4; DB 7; Length 3587;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCGCACACCTCGACTT 20
DB 3003 GCAGCACCACCTCGACTT 3020

RESULT 28
RAD12302
ID AAD12302 standard; cDNA; 3600 BP.
XX AC AAD12302;
XX XX
XX DT 16-OCT-2001 (first entry)
XX DE Human Cubitus interruptus (Ci) homologue, GLI-1 cDNA.

XX KW Human; transgenic non-human animal; Cubitus interruptus; Ci; GLI-1;
KW basal cell carcinoma; BCC model system; tumour; screening; anti-cancer;
KW trichioepithelioma; cylindroma; trichoblastoma; ss.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 79..3399
XX FT /*tag= a
XX FT /product= "Human Ci homologue, GLI-1"
XX XX
XX PN WO200156376-A1.
XX PD 09-AUG-2001.
XX XX
XX PF 02-FEB-2001; 2001WO-SE000204.
XX XX
XX PR 03-FEB-2000; 2000SE-00000345.
XX XX
XX PA (KARO-) KAROLINSKA INNOVATIONS AB.
XX PI Toftgard R;
XX DR WPI: 2001-488828/53.
XX DR P-PSDB; AA306644.
XX XX
XX PT Transgenic non-human animal useful as basal cell carcinoma model system
XX PT to identify anti-cancer drug candidates, overexpresses transgene encoding
XX PT GLI-1 protein which is a human homolog to Cubitus interruptus.
XX PS Claim 6; Page 25-26; 33pp; English.
XX CC The present invention relates to a transgenic non-human animal comprising
XX CC a transgene containing a nucleic acid encoding a human Cubitus
XX CC interruptus (Ci) homologue protein, GLI-1. The transgenic non-human
XX CC animal is useful as basal cell carcinoma (BCC) model system since it
XX CC overexpresses GLI-1 which leads to development of tumours resembling
XX CC human BCC. Thus it is also useful for screening anti-cancer drug
XX CC candidates and evaluating whether it affects BCC, trichioepitheliomas,
XX CC cylindromas and trichoblastomas. The present sequence is a cDNA encoding
XX CC human Ci homologue protein, GLI-1
XX SQ Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 4; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCCGCACACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051

RESULT 29
ABK30501
ID ABK30501 standard; DNA; 3600 BP.
XX AC ABK30501;
XX XX
XX DT 23-APR-2002 (first entry)
XX DE Human glioma-associated oncogene-1 DNA sequence.
XX KW Human; glioma-associated oncogene-1 associated disease; infection;
XX KW inflammation; tumour formation; cytostatic; antiinflammatory; gene; ds.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 79..3399
XX FT /*tag= a
XX FT /product= "Glioma-associated oncogene-1 protein"

XX US6329203-B1.
PN
XX 11-DEC-2001.
PD
XX 08-SEP-2000; 2000US-00657042.
PP
XX 08-SEP-2000; 2000US-00657042.
PR
XX (ISIS-) ISIS PHARM INC.
PA
XX Bennett CF, Wyatt J;
PI
XX WPI; 2002-138363/18.
XX
DR P-PSDB; AAU12085.
DR
XX Novel antisense compounds targeted to nucleic acids encoding glioma-
PT associated oncogene-1, for modulating the gene expression and treating
PT diseases associated with expression of the oncogene in humans.
PT
XX Example 13; Col 47-56; 43pp; English.
PS
XX The present invention relates to antisense compounds and methods for
CC modulating the expression of human glioma-associated oncogene-1. The
CC antisense compounds, particularly antisense oligonucleotides (ABK30509-
CC ABK30586), target and inhibit the expression of human glioma-associated
CC oncogene-1. The antisense compounds are useful for inhibiting the
CC expression of human glioma-associated oncogene-1 in human cells or
CC tissues and for treating an animal, particularly a human suspected of
CC having or being prone to a disease or condition associated with
CC expression of glioma-associated oncogene-1. The compounds are useful for
CC diagnostics, therapeutics and as research reagent, e.g. prophylactically
CC to prevent or delay infection, inflammation or tumour formation. The
CC antisense compounds are safely and effectively administered to humans.
CC The present sequence represents human glioma-associated oncogene-1 DNA
XX
SQ Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 6; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGACCACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051

RESULT 30
AAL45544
ID AAL45544 standard; cDNA; 3600 BP.
XX
XX AAL45544;
AC
XX 11-JUN-2002 (first entry)
DT
XX Human Gli1 coding sequence SEQ ID NO: 16.
DE
XX Gli1; screening method; bone induction; cartilage induction;
KW orthopaedic disease; dental disease; osteoporosis; hyperosteoecgenesis;
KW osteopathic; antiarthritic; vulnary; immunosuppressive; human;
KW hyperchondrogenesis; gene; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 79..3399
FT CDS /*tag= a
FT /product= "Gli1"
FT /transl_except= (pos:2875..2877,aa:Glu)
XX
PN WO200211752-A1.
XX
PD 14-FEB-2002.

XX 03-AUG-2001; 2001WO-JF006688.
PF
XX 04-AUG-2000; 2000JP-00242767.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Hikichi Y;
PI
XX WPI; 2002-241709/29.
XX
DR P-PSDB; AAO17111.
DR
XX Promotion of bone and cartilage formation using Gli1 protein or DNA
PT encoding it for treatment of skeletal disorders.
PT
XX Claim 6; Page 115-117; 154pp; Japanese.
PS
XX The present invention relates to agents for the promotion of bone and
CC cartilage formation which contain as the active component a Gli1 protein
CC or a DNA encoding a Gli1 protein. The agents can be used in the
CC prevention, treatment and diagnosis of bone and cartilage disorders
CC including bone fractures, joint deformation, osteoarthritis,
CC osteoporosis, cartilage damage, trauma, bone formation defects, cartilage
CC formation defects, bone defects, dental disease, hyperosteoecgenesis and
CC hyperchondrogenesis, and for use in cosmetic and therapeutic bone
CC transplantation. The present sequence is a human Gli1 coding sequence
CC described in the exemplification of the invention
XX
SQ Sequence 3600 BP; 786 A; 1161 C; 948 G; 705 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 6; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGACCACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051

Search completed: June 20, 2004, 10:17:02
Job time : 133.046 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 934.603 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 gcgcgcacacacacgcactt 20

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 300 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	1040	29	AG056593 Pan trogl
2	16.8	84.0	223	12	BI470080 saf38ell.
3	16.8	84.0	298	12	BI239271 RE35720.5
4	16.8	84.0	305	13	BQ102855 UUGC0247

5	16.8	84.0	309	12	BI212140
6	16.8	84.0	310	12	BI172928
7	16.8	84.0	333	10	BF486064
8	16.8	84.0	333	12	BI241833
9	16.8	84.0	335	12	BI363139
10	16.8	84.0	336	12	BI240825
11	16.8	84.0	345	12	BI578091
12	16.8	84.0	346	12	BI227580
13	16.8	84.0	356	12	BI361366
14	16.8	84.0	364	9	AI061909
15	16.8	84.0	378	9	AA201741
16	16.8	84.0	393	12	BI243125
17	16.8	84.0	397	12	BI369591
18	16.8	84.0	399	12	BI239432
19	16.8	84.0	399	12	BI356076
20	16.8	84.0	404	12	BI357752
21	16.8	84.0	414	9	AA695426
22	16.8	84.0	414	12	BI577942
23	16.8	84.0	418	13	BQ103130
24	16.8	84.0	419	13	BQ103552
25	16.8	84.0	427	12	BI163420
26	16.8	84.0	427	12	BI213953
27	16.8	84.0	427	12	BI375669
28	16.8	84.0	427	12	BI484865
29	16.8	84.0	428	12	BI365791
30	16.8	84.0	428	12	BI366732
31	16.8	84.0	429	12	BI239241
32	16.8	84.0	430	12	BI213411
33	16.8	84.0	441	12	BI165466
34	16.8	84.0	448	12	BI212897
35	16.8	84.0	452	12	BI239186
36	16.8	84.0	453	12	BI240875
37	16.8	84.0	454	9	AA201722
38	16.8	84.0	456	12	BI375564
39	16.8	84.0	457	12	BI212874
40	16.8	84.0	457	12	BI363271
41	16.8	84.0	458	12	BI481967
42	16.8	84.0	460	12	BI608622
43	16.8	84.0	461	12	BI241749
44	16.8	84.0	462	12	BI370712
45	16.8	84.0	465	12	BI164419
46	16.8	84.0	467	9	AI135940
47	16.8	84.0	467	9	AA201915
48	16.8	84.0	467	12	BI582929
49	16.8	84.0	468	12	BI486175
50	16.8	84.0	475	12	BI608870
51	16.8	84.0	476	12	BI167792
52	16.8	84.0	476	12	BI579620
53	16.8	84.0	476	12	BI581169
54	16.8	84.0	477	10	BF501960
55	16.8	84.0	480	12	BI172447
56	16.8	84.0	482	12	BI228973
57	16.8	84.0	483	10	BF419339
58	16.8	84.0	484	9	AI296025
59	16.8	84.0	486	12	BI356678
60	16.8	84.0	486	12	BI372713
61	16.8	84.0	487	12	BI356519
62	16.8	84.0	487	12	BI362624
63	16.8	84.0	489	9	AA695778
64	16.8	84.0	492	12	BI171646
65	16.8	84.0	492	12	BI365023
66	16.8	84.0	493	12	BI171638
67	16.8	84.0	493	12	BI605905
68	16.8	84.0	494	12	BI363780
69	16.8	84.0	501	12	BI593455
70	16.8	84.0	501	13	BQ103380
71	16.8	84.0	501	13	BI627342
72	16.8	84.0	503	12	BI584671
73	16.8	84.0	503	12	BI627199
74	16.8	84.0	504	29	TA226802Q
75	16.8	84.0	506	12	BI581210
76	16.8	84.0	507	12	BI608523
77	16.8	84.0	515	12	BI170929

C

78	16.8	84.0	516	12	BI213717	RE19185.5	151	16.8	84.0	590	9	AI519385
79	16.8	84.0	516	12	BI214879	RE21025.5	152	16.8	84.0	590	10	BE585438
80	16.8	84.0	516	12	BI368926	RE54986.5	153	16.8	84.0	591	9	AA438350
81	16.8	84.0	516	29	CG233095	OGVC891TV	154	16.8	84.0	591	12	BI574950
82	16.8	84.0	518	12	BI613418	RE42567.5	155	16.8	84.0	591	12	BI574950
83	16.8	84.0	519	12	BI618036	RE48552.5	156	16.8	84.0	592	9	AI589287
84	16.8	84.0	520	12	BI611182	RE17337.5	157	16.8	84.0	592	9	AA390375
85	16.8	84.0	525	9	AI546641	SD09804.5	158	16.8	84.0	592	12	BI565002
86	16.8	84.0	526	12	BI368300	RE54249.5	159	16.8	84.0	593	12	BI364594
87	16.8	84.0	527	12	BI364633	RE49794.5	160	16.8	84.0	593	12	BI365916
88	16.8	84.0	528	12	BI580722	RE74945.5	161	16.8	84.0	594	12	BI615292
89	16.8	84.0	528	9	AI294617	LP08032.5	162	16.8	84.0	594	12	BI617933
90	16.8	84.0	528	12	BI168132	RE08688.5	163	16.8	84.0	595	12	BI629243
91	16.8	84.0	528	12	BI170588	RE12104.5	164	16.8	84.0	596	12	BI213057
92	16.8	84.0	528	12	BI170656	RE12181.5	165	16.8	84.0	596	12	BI482586
93	16.8	84.0	528	12	BI173768	RE16967.5	166	16.8	84.0	597	12	BI627129
94	16.8	84.0	528	12	BI228008	RE25417.5	167	16.8	84.0	597	12	BI241644
95	16.8	84.0	528	12	BI229663	RE27490.5	168	16.8	84.0	599	12	BI371839
96	16.8	84.0	528	12	BI234104	RE29745.5	169	16.8	84.0	599	12	BI365336
97	16.8	84.0	528	12	BI235906	RE31860.5	170	16.8	84.0	599	12	BI366350
98	16.8	84.0	528	12	BI365136	RE50367.5	171	16.8	84.0	599	28	AI2300928
99	16.8	84.0	528	12	BI365546	RE52109.5	172	16.8	84.0	600	12	BI168852
100	16.8	84.0	528	12	BI365744	RE52345.5	173	16.8	84.0	600	12	BI229142
101	16.8	84.0	528	12	BI367407	RE53688.5	174	16.8	84.0	600	12	BI229142
102	16.8	84.0	528	12	BI374213	RE61747.5	175	16.8	84.0	600	12	BI364207
103	16.8	84.0	529	12	BI164845	RE04596.5	176	16.8	84.0	600	12	BI364331
104	16.8	84.0	529	12	BI230985	RE16659.5	177	16.8	84.0	601	12	BI578827
105	16.8	84.0	529	12	BI358430	RE45196.5	178	16.8	84.0	601	12	BI577743
106	16.8	84.0	531	12	BI215834	RE22567.5	179	16.8	84.0	602	9	AI259223
107	16.8	84.0	531	12	BI361983	RE26430.5	180	16.8	84.0	602	12	BI161506
108	16.8	84.0	532	12	BI370861	RE57829.5	181	16.8	84.0	602	12	BI235077
109	16.8	84.0	534	12	BI368975	RE55044.5	182	16.8	84.0	602	12	BI365077
110	16.8	84.0	535	12	BI568993	RE01228.5	183	16.8	84.0	604	9	AA440041
111	16.8	84.0	537	12	BI484987	RE68192.5	184	16.8	84.0	604	12	BI164661
112	16.8	84.0	538	14	CP861330	P8200071B	185	16.8	84.0	604	12	BI375713
113	16.8	84.0	539	12	BI572338	RE34659.5	186	16.8	84.0	605	12	BI235202
114	16.8	84.0	546	12	BI363769	RE48710.5	187	16.8	84.0	606	12	BI172172
115	16.8	84.0	547	12	BI241414	RE38472.5	188	16.8	84.0	606	12	BI591997
116	16.8	84.0	547	12	BI365535	RE42711.5	189	16.8	84.0	606	12	BI591997
117	16.8	84.0	548	12	BI370055	RE56431.5	190	16.8	84.0	606	12	BI6025603
118	16.8	84.0	548	12	BI481540	RE64111.5	191	16.8	84.0	606	12	BI624943
119	16.8	84.0	556	12	BI370600	RE57293.5	192	16.8	84.0	607	12	BI169813
120	16.8	84.0	557	12	BI243961	RE41783.5	193	16.8	84.0	608	9	AA735381
121	16.8	84.0	559	12	BI375475	RE63302.5	194	16.8	84.0	608	9	AI108250
122	16.8	84.0	560	12	BI366330	RE51841.5	195	16.8	84.0	608	12	BI584497
123	16.8	84.0	561	12	BI242099	RE39328.5	196	16.8	84.0	608	14	CA807327
124	16.8	84.0	561	12	BI243597	RE41384.5	197	16.8	84.0	609	10	BE977127
125	16.8	84.0	562	12	BI216600	RE23794.5	198	16.8	84.0	609	12	BI232434
126	16.8	84.0	562	12	BI367617	RE53419.5	199	16.8	84.0	610	12	BI364611
127	16.8	84.0	562	12	BI486118	RE69503.5	200	16.8	84.0	610	12	BI215547
128	16.8	84.0	563	9	AI390531	LD08747.5	201	16.8	84.0	610	12	BI568122
129	16.8	84.0	563	12	BI240070	RE36717.5	202	16.8	84.0	610	12	BI607657
130	16.8	84.0	563	12	BI486757	RE70221.5	203	16.8	84.0	611	12	BI607657
131	16.8	84.0	564	12	BI617338	RE46078.5	204	16.8	84.0	611	12	BI615328
132	16.8	84.0	565	12	BI163984	RE30335.5	205	16.8	84.0	611	12	BI604848
133	16.8	84.0	565	12	BI368351	RE254307.5	206	16.8	84.0	613	12	BI626443
134	16.8	84.0	566	9	AI539800	LD19053.5	207	16.8	84.0	614	12	BI165524
135	16.8	84.0	568	12	BI366714	RE52310.5	208	16.8	84.0	614	12	AI517455
136	16.8	84.0	568	12	BI579198	RE73188.5	209	16.8	84.0	614	12	BI172007
137	16.8	84.0	570	9	AI389199	GH20316.5	210	16.8	84.0	614	12	BI358429
138	16.8	84.0	571	12	BI576684	RE44221.5	211	16.8	84.0	615	14	CA805518
139	16.8	84.0	572	12	BI162252	RE01620.5	212	16.8	84.0	616	12	BI357706
140	16.8	84.0	573	12	BI618133	RE48692.5	213	16.8	84.0	616	12	BI579025
141	16.8	84.0	577	9	AI238765	GH14872.5	214	16.8	84.0	617	12	BI164335
142	16.8	84.0	577	12	BI238504	RE34985.5	215	16.8	84.0	617	12	BI231890
143	16.8	84.0	577	12	BI236356	RE49417.5	216	16.8	84.0	617	12	BI373623
144	16.8	84.0	577	29	CG624831	OGXACS0TV	217	16.8	84.0	617	12	BI375535
145	16.8	84.0	578	9	AI389734	GH20975.5	218	16.8	84.0	618	9	AI110215
146	16.8	84.0	578	12	BI241945	RE39151.5	219	16.8	84.0	618	12	BI367381
147	16.8	84.0	578	12	BI623417	RE55662.5	220	16.8	84.0	619	12	BI375379
148	16.8	84.0	583	12	BI167321	RE07678.5	221	16.8	84.0	620	9	AI238530
149	16.8	84.0	585	12	BI214092	RE19802.5	222	16.8	84.0	620	9	AI258403
150	16.8	84.0	586	12	BI364504	RE49338.5	223	16.8	84.0	621	12	BI215987
151	16.8	84.0	586	12	BI364504	RE49338.5	223	16.8	84.0	621	12	BI215987

```
224 16.8 84.0 621 12 BI565840
225 16.8 84.0 623 12 BI616671
226 16.8 84.0 624 9 AI297684
227 16.8 84.0 624 12 BI364910
228 16.8 84.0 625 12 BI591996
229 16.8 84.0 626 9 AI405628
230 16.8 84.0 630 9 AI295441
231 16.8 84.0 630 10 BF498793
232 16.8 84.0 630 12 BI615419
233 16.8 84.0 630 29 CC752514
234 16.8 84.0 631 12 BI232468
235 16.8 84.0 631 12 BI580703
236 16.8 84.0 631 12 BI580810
237 16.8 84.0 632 12 BI363330
238 16.8 84.0 633 12 BI169213
239 16.8 84.0 633 12 BI372162
240 16.8 84.0 633 29 AG116253
241 16.8 84.0 634 12 BI214904
242 16.8 84.0 634 12 BI242789
243 16.8 84.0 634 12 BI629921
244 16.8 84.0 635 12 BI228617
245 16.8 84.0 635 12 BI565430
246 16.8 84.0 635 12 BI573264
247 16.8 84.0 636 12 BI169784
248 16.8 84.0 637 12 BI228554
249 16.8 84.0 637 12 BI371288
250 16.8 84.0 638 12 BI484890
251 16.8 84.0 639 9 AI238918
252 16.8 84.0 639 9 AI519051
253 16.8 84.0 639 12 BI629652
254 16.8 84.0 641 12 BI579153
255 16.8 84.0 641 12 BI365208
256 16.8 84.0 641 13 BQ523510
257 16.8 84.0 643 9 AA441085
258 16.8 84.0 644 12 BI579699
259 16.8 84.0 646 9 AI388592
260 16.8 84.0 646 12 BI237780
261 16.8 84.0 646 12 BI487106
262 16.8 84.0 646 12 BI577177
263 16.8 84.0 646 12 BI579281
264 16.8 84.0 647 9 AA736185
265 16.8 84.0 647 12 BI229262
266 16.8 84.0 647 12 BI371214
267 16.8 84.0 648 12 BI169408
268 16.8 84.0 648 12 BI216098
269 16.8 84.0 648 12 BI227540
270 16.8 84.0 648 12 BI238417
271 16.8 84.0 649 12 BI169356
272 16.8 84.0 650 9 AI063600
273 16.8 84.0 650 12 BI216602
274 16.8 84.0 650 12 BI236631
275 16.8 84.0 650 14 CA803177
276 16.8 84.0 651 12 BI213937
277 16.8 84.0 652 9 AA696210
278 16.8 84.0 652 9 AA942427
279 16.8 84.0 652 12 BI216725
280 16.8 84.0 654 12 BI375477
281 16.8 84.0 656 9 AI386508
282 16.8 84.0 656 12 BI231861
283 16.8 84.0 656 12 BI625747
284 16.8 84.0 657 12 BI234575
285 16.8 84.0 657 12 BI236602
286 16.8 84.0 657 12 BI239825
287 16.8 84.0 657 12 BI580105
288 16.8 84.0 658 12 BI228473
289 16.8 84.0 658 12 BI577865
290 16.8 84.0 658 12 BI586497
291 16.8 84.0 658 12 BI638518
292 16.8 84.0 659 12 BI167175
293 16.8 84.0 660 12 BI231162
294 16.8 84.0 661 12 BI238676
295 16.8 84.0 662 12 BI238862
296 16.8 84.0 662 12 BI372915

BI565840 RH63667.5
BI616671 RH46718.5
AI297684 LP12044.5
BI364910 RE50110.5
BI591996 RH09577.5
AI405628 GH25714.5
AI295441 LP09068.5
BF498793 AT13265.5
BI615419 RH45040.5
CC752514 ZMREB013
BI232468 RE28539.5
BI580703 RE74926.5
BI580810 RE75047.5
BI363330 RE48110.5
BI169213 RE10323.5
BI372162 RE59049.5
AG116253 Pan trogl
BI214904 RE21054.5
BI242789 RE40217.5
BI629921 RH58860.5
BI228617 RE26214.5
BI565430 RH63205.5
BI573264 RH09957.5
BI169784 RE11056.5
BI228554 RE26126.5
BI371288 RE58152.5
BI484890 RE68064.5
AI238918 GH15057.5
AI519051 LD31875.5
BI629652 RH58594.5
BI579153 RE73139.5
BI365208 RE50450.5
BQ523510 NISC.n122
AA441085 LD15868.5
BI579699 RE73769.5
AI388592 GH19549.5
BI237780 RE34073.5
BI487106 RE70616.5
BI577177 RE70827.5
BI579281 RE73281.5
AA736185 HL08132.5
BI229262 RE27012.5
BI371214 RE58061.5
BI169408 RE10575.5
BI216098 RE23014.5
BI227540 RE24618.5
BI238417 RE34873.5
BI169356 RE10506.5
AI063600 GH03506.5
BI216602 RE23796.5
BI236631 RE32710.5
CA803177 ESG011A.E
BI213937 RE19586.5
AA696210 GH05286.5
AA942427 LD26644.5
BI216725 RE23956.5
BI375477 RE63305.5
AI386508 GH16608.5
BI231861 RE24807.5
BI625747 RH66231.5
BI234575 RE30293.5
BI236602 RE32674.5
BI239825 RE36422.5
BI580105 RE74242.5
BI228473 RE26026.5
BI577865 RE71623.5
BI586497 RH26453.5
BI638518 SD20821.5
BI167175 RE207504.5
BI231162 RE19376.5
BI238676 RE35184.5
BI238862 RE41987.5
BI372915 RE60114.5

RESULT 1
AG056593 1040 bp DNA linear GSS 02-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-042020.F, genomic survey sequence.
DEFINITION AG056593
ACCESSION AG056593
VERSION AG056593.1 GI:16594052
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1040)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.sc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKSI45
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
1..1040
location/Qualifiers
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-042020.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match 85.0%; Score 17; DB 29; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCGCGCACCACCTCGA 17
DB 372 GCGCGCACCACCTCGA 388
RESULT 2
BI470080/c 223 bp mRNA linear EST 29-NOV-2001
LOCUS saf38cell.y3 Gm-cl077 Glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-cl077-981.5, similar to TR:Q9T0L6 Q9T0L6 HYPOTHETICAL 23.2
KD PROTEIN. ;, mRNA sequence.
ACCESSION BI470080
VERSION BI470080.1 GI:15286189
KEYWORDS EST.
SOURCE Glycine max (soybean)
```


SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 305)
 AUTHORS Simin, K., Scuderi, A., Reamey, J., Dunn, D.M., Weiss, R.B., Metherall, J.E. and Letsou, A.
 TITLE An automated screen for spatially restricted transcripts in Drosophila embryogenesis
 JOURNAL Genome Res. (2002) In press
 COMMENT Contact: Letsou, A.
 Department of Human Genetics
 Eccles Institute of Human Genetics, University of Utah
 15 N 2030 EAST RM 2100, Salt Lake City, UT 84112-5330, USA
 Tel: 801 581-4422
 Fax: 801 581-7796
 Email: aletsou@genetics.utah.edu
 Seq primer: SP6
 Location/Qualifiers
 FEATURES
 source 1..305
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /dev_stage="embryonic 8-12 hr post-fertilization"
 /clone_lib="Drosophila 8-12 hr embryonic cDNA library"
 /note="Vector: plasmid pNB; see Brown, N.H. and Kafatos, F.C. (1988) Functional cDNA libraries from Drosophila embryos. J. Mol Biol. 203, 425-37."
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 13; Length 305;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCCTCGACTT 20
 |||||
 Db 240 GCTCCGCCACCTCGACTT 259
 RESULT 5
 B1212140 309 bp mRNA linear EST 11-JUL-2001
 LOCUS RE17164.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
 DEFINITION Drosophila melanogaster cDNA clone RE17164 5 similar to guf:
 Fban0016747 'enzyme inhibitor' located on: 2R 49F1-48F1;:
 04/12/2001, mRNA sequence.
 ACCESSION B1212140 GI:14689864
 VERSION B1212140.1
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 309)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
 TITLE BDGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AEO03823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 04/12/2001
 Plate: RE.171 row: F column: 4
 High quality sequence stop: 4
 Location/Qualifiers
 FEATURES
 source 1..310
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE15633"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha ToxA"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

High quality sequence stop: 272.
 Location/Qualifiers
 FEATURES
 source 1..309
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE17164"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha ToxA"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 309;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCCTCGACTT 20
 |||||
 Db 238 GCTCCGCCACCTCGACTT 257
 RESULT 6
 B1172928 310 bp mRNA linear EST 09-JUL-2001
 LOCUS RE15633.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
 DEFINITION Drosophila melanogaster cDNA clone RE15633 5 similar to guf:
 Fban0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;:
 04/12/2001, mRNA sequence.
 ACCESSION B1172928 GI:14638735
 VERSION B1172928.1
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 310)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
 TITLE BDGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AEO03823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 04/12/2001
 Plate: RE.156 row: C column: 9
 High quality sequence stop: 276.
 Location/Qualifiers
 FEATURES
 source 1..310
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE15633"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha ToxA"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 310;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
DB 238 GCTCCGCACCACTCGACTT 257

RESULT 7

BF486064 333 bp mRNA linear EST 19-APR-2001
LOCUS AT20210.5prime AT Drosophila melanogaster adult testes pOT57
DEFINITION Drosophila melanogaster cDNA clone AT20210.5 similar to gulf.
FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
04/09/2001, mRNA sequence.

ACCESSION

BF486064

VERSION

EST

KEYWORDS

EST

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 333)

REFERENCE

AUTHORS

Stapleton, M., Brockstein, P., Hong, L., Aghayani, A., Baxter, E.,

Berman, B., Carlson, J., Champe, M., Chavez, C., Chev, M., Dorsett, V.,

Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,

Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunco, J.,

Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,

Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11569365.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu

hit genomic AE003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 04/09/2001

Plate: AT.202 row: A column: 10

High quality sequence stop: 324.

Location/Qualifiers

1..333

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clones="AT20210"

/sex="male"

/dev_stages="0-3 day old Ore-R males"

/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates

AT.121-AT.319: DH5-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes pOT57"

/note="Organ: ADULT testes; Vector: pOT57; Site:1: EcoRI;

Site 2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into pOT57. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 333;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20

DB 117 GCTCCGCACCACTCGACTT 136

RESULT 8

BI241833

LOCUS

DEFINITION

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 333)

Stapleton, M., Brockstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D.,

Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,

Liao, G., Misra, S., Mungall, C.J., Nunco, J., Pacleb, J.,

Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,

Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu

hit genomic AE003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 05/12/2001

Plate: RE.390 row: B column: 4

High quality sequence stop: 306.

Location/Qualifiers

1..333

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clones="RE39016"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo

pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 333;

Best Local Similarity 90.0%; Pred. No. 1.5e+04;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20

DB 238 GCTCCGCACCACTCGACTT 257

RESULT 9

BI363139

LOCUS

DEFINITION

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 333)

Stapleton, M., Brockstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D.,

Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,

Liao, G., Misra, S., Mungall, C.J., Nunco, J., Pacleb, J.,

Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,

Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu

hit genomic AE003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 05/13/2001, mRNA sequence.

Plate: RE.390 row: B column: 4

High quality sequence stop: 306.

Location/Qualifiers

1..333

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clones="RE39016"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo

pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

SOURCE
ORGANISM Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS 1 (bases 1 to 335)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AEO03823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 05/13/2001
 Plate: RE.478 row: G column: 8
 High quality sequence stop: 334.
FEATURES
 Location/Qualifiers
 source
 1..335
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE47880"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 335;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||||
 Db 129 GCTCCGCACCATCTCGACTT 148
RESULT 10
LOCUS B1240825 336 bp mRNA linear EST 12-JUL-2001
DEFINITION RE37759.Sprime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE37759 5 similar to gtf:
 Fban0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
 05/12/2001, mRNA sequence.
ACCESSION B1240825
VERSION B1240825.1 GI:14709461
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 336)
REFERENCE
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL
COMMENT Unpublished (2001)
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AEO03823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 05/12/2001
 Plate: RE.377 row: E column: 11
 High quality sequence stop: 288.
FEATURES
 Location/Qualifiers
 source
 1..336
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE37759"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 336;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||||
 Db 130 GCTCCGCACCATCTCGACTT 149
RESULT 11
LOCUS B1578091 345 bp mRNA linear EST 06-SEP-2001
DEFINITION RE71891.Sprime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE71891 5 similar to gtf:
 Fban0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
 located on: 2R 48F1-48F1; 05/18/2001, mRNA sequence.
ACCESSION B1578091
VERSION B1578091.1 GI:15469513
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 345)
REFERENCE
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: RE.718 row: H column: 7
 High quality sequence stop: 292.
FEATURES
 Location/Qualifiers
 source
 1..345
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"

/db_xref="taxon:7227"
 /clone="RE71891"
 /sex="male and female"
 /dev_stages="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 15.8; DB 12; Length 345;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 238 GCTCCGCACCATCTCGACTT 257

RESULT 12

B1227580

LOCUS

DEFINITION B1227580 346 bp mRNA linear EST 11-JUL-2001
 RE24669.Sprime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE24669 5 similar to guf.
 FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1.1;

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 346)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
 Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT

Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AB003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 04/12/2001

Plate: RE.246 row: F column: 9

High quality sequence stop: 276.

FEATURES

source

1..346
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE24669"
 /sex="male and female"
 /dev_stages="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 346;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 238 GCTCCGCACCATCTCGACTT 257

RESULT 13

B1361366

LOCUS

DEFINITION B1361366 356 bp mRNA linear EST 01-AUG-2001
 RE45550.Sprime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE45550 5 similar to guf.
 FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1.1;

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 356)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
 Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT

Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AB003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 05/13/2001

Plate: RE.455 row: E column: 2

High quality sequence stop: 311.

FEATURES

source

1..356
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE45550"
 /sex="male and female"
 /dev_stages="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 356;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 130 GCTCCGCACCATCTCGACTT 149

RESULT 14

A1061909

LOCUS

DEFINITION A1061909 354 bp mRNA linear EST 19-APR-2001
 LD34973.Sprime LD Drosophila melanogaster embryo pOT2 Drosophila

melanogaster cDNA clone LD34973 5prime similar to AF038597;
Drosophila melanogaster ornithine decarboxylase antizyme mRNA,
complete cds, mRNA sequence.

ACCESSION AT061909.1 GI:3337748

VERSION EST

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 364)

AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G.,

LEWIS, S. and Rubin, G.M.

TITLE BGP/HMI Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

Plate: 349 row: G column: 1

High quality sequence stop: 352.

FEATURES

Location/Qualifiers

1..364

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LD34973"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="X11 Blue"

/clone_lib="LD Drosophila melanogaster embryo pOR2"

/note="Organ: embryo; Vector: pOR2; Site: 1; EcoRI; Site: 2;

XhoI; Sized fractionated cDNAs were directly ligated into

pOR2."

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 364;

Best Local Similarity 90.0%; Pred. No. 1.5e+04;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGCGCACCATCTCGACTT 20

|||||

63 GCTCCGCACCATCTCGACTT 82

RESULT 15

AA201741

LOCUS

DEFINITION LD04952.5prime LD Drosophila melanogaster embryo BlueScript

Drosophila melanogaster cDNA clone LD04952 5prime, mRNA sequence.

ACCESSION AA201741

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 378)

AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G.,

LEWIS, S. and Rubin, G.M.

TITLE BGP/HMI Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1797526.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

Blast expect value = 5.0E-87 on U29529: Drosophila melanogaster
gutfeeling (guf) mRNA, complete cds
Plate: 49 row: E column: 4
High quality sequence stop: 305.

FEATURES

Location/Qualifiers

1..378

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="BDGP_EST:BDcln004335"

/db_xref="taxon:7227"

/clone="LD04952"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="SOLR"

/clone_lib="LD Drosophila melanogaster embryo Bluescript"

/note="Organ: embryo; Vector: Bluescript SK; Site: 1;

EcoRI; Site: 2; XhoI; Constructed using Stratagene ZAP-cDNA

Synthesis kit. Oligo dt-primed and directionally cloned at

EcoRI and XhoI in Bluescript SK(+/-)"

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 378;

Best Local Similarity 90.0%; Pred. No. 1.5e+04;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGCGCACCATCTCGACTT 20

|||||

201 GTCGCGCACCATCTCGACTT 220

RESULT 16

BI243125

LOCUS

DEFINITION

BI243125

393 bp mRNA linear EST 12-JUL-2001

RE40682.5prime RE Drosophila melanogaster normalized Embryo pFLC-1

Drosophila melanogaster cDNA clone RE40682 5 similar to guf;

FBanc006747; enzyme inhibitor; located on: 2R 48F1-48F1.1;

05/12/2001, mRNA sequence.

ACCESSION BI243125

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 393)

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

Misra, S., Mungall, C.J., Nunco, J., Pacleb, J., Paragas, V., Park, S.,

Phuanavanvorong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and

Rubin, G.M.

TITLE BGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AE003823; arm: 2R [697351, 7260081]

estimated-cyto: 48D8-48F5: 05/12/2001

Plate: RE:406 row: G column: 10

High quality sequence stop: 322.

FEATURES

Location/Qualifiers

1..393

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE40682"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 399)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AB003823; arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5; 05/13/2001
 Plate: RE.421 row: B column: 2
 High quality sequence stop: 346.
 Location/Qualifiers
 1. 399
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="R842114"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
FEATURES
 source
 1. 399
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="R842114"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 399;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GCGCGCGCACCACTCGACTT 20
 |||||||
 Db 238 GCTCGGACCATCTCGACTT 257
 |||||||
RESULT 20
BI357752 404 bp mRNA linear EST 31-JUL-2001
LOCUS R844306.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
DEFINITION Drosophila melanogaster cDNA clone R844306 5 similar to guf;
 F8an0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
 05/13/2001, mRNA sequence.
ACCESSION BI357752
VERSION BI357752.1 GI:15052206
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 404)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AB003823; arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5; 05/13/2001
 Plate: RE.443 row: A column: 6
 High quality sequence stop: 292.
 Location/Qualifiers
 1. 404
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE44306"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 404;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GCGCGCGCACCACTCGACTT 20
 |||||||
 Db 237 GTCGCGACCATCTCGACTT 256
 |||||||
RESULT 21
AA695426 414 bp mRNA linear EST 23-APR-2001
LOCUS GM02813.5prime GM Drosophila melanogaster ovary Bluescript
DEFINITION Drosophila melanogaster cDNA clone GM02813 5prime, mRNA sequence.
ACCESSION AA695426
VERSION AA695426.1 GI:2698046
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 414)
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: 28 row: B column: 1
 High quality sequence stop: 206.
 Location/Qualifiers
 1. 414
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="GM02813"
 /sex="female"
 /dev_stage="newly enclosed females: germarium-stage 6"
 /lab_host="SOLR"

/clone lib="GM Drosophila melanogaster ovary Bluescript"
 /note="organ: ovary; Vector: Bluescript SK; Site 1: EcoRI;
 Site 2: XhoI; Constructed using Stratagene ZAP-cDNA
 Synthesis kit. Oligo dt-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 414;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 212 GCTCCGCACCACTCGACTT 231

RESULT 22

B1577942 414 bp mRNA linear EST 06-SEP-2001
 RE71718.5prime RE Drosophila melanogaster normalized Embryo pf1c-1
 Drosophila melanogaster cDNA clone RE71718 5 similar to gtf:
 FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
 located on: 2R 48F1-48F1; : 05/18/2001, mRNA sequence.

ACCESSION B1577942.1 GI:15469364
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 414)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C. J., Nuno, J., Pacle, J., Paragas, V., Park, S.,
 Phuanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and
 Rubin, G. M.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

COMMENT

BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estofruitfly.berkeley.edu
 hit genomic A5003823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 05/18/2001
 Plate: RE.717 row: B column: 6
 High quality sequence stop: 308.

FEATURES

Location/Qualifiers
 1..414
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE71718"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha Tonk"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pf1c-1"

/note="Organ: embryo; Vector: pf1c1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 414;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||

Db 238 GCTCCGCACCACTCGACTT 257

RESULT 23

BQ103130 418 bp mRNA linear EST 12-APR-2002
 LOCUS
 DEFINITION
 UUGC0528 Drosophila 8-12 hr embryonic cDNA library Drosophila
 melanogaster cDNA 5', mRNA sequence.

ACCESSION BQ103130
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 418)
 Simin, K., Scuderi, A., Reamey, J., Dunn, D. M., Weiss, R. B.,
 Metherall, J. E. and Letsou, A.
 An automated screen for spatially restricted transcripts in
 Drosophila embryogenesis
 Genome Res. (2002) In press
 Contact: Letsou, A.

JOURNAL

COMMENT
 Department of Human Genetics
 Eccles Institute of Human Genetics, University of Utah
 15 N 2030 EAST RM 2100, Salt Lake City, UT 84112-5330, USA
 Tel: 801 581-4422
 Fax: 801 581-7796
 Email: aletsou@genetics.utah.edu

Seq primer: SP6.

FEATURES

Location/Qualifiers
 1..418
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /dev_stage="embryonic 8-12 hr post-fertilization"
 /clone_lib="Drosophila 8-12 hr embryonic cDNA library"
 /note="Vector: plasmid pNB; see Brown, N.H. and Kafatos,
 F.C. (1988) Functional cDNA libraries from Drosophila
 embryos. J. Mol Biol. 203, 425-37."

ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 418;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||

Db 194 GCTCCGCACCACTCGACTT 213

RESULT 24

BQ103552 419 bp mRNA linear EST 12-APR-2002
 LOCUS
 DEFINITION
 UUGC0559 Drosophila 8-12 hr embryonic cDNA library Drosophila
 melanogaster cDNA 5', mRNA sequence.

ACCESSION BQ103552
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 419)
 Simin, K., Scuderi, A., Reamey, J., Dunn, D. M., Weiss, R. B.,
 Metherall, J. E. and Letsou, A.

An automated screen for spatially restricted transcripts in

Drosophila embryogenesis

Genome Res. (2002) In press

JOURNAL

COMMENT
 Contact: Letsou, A.
 Department of Human Genetics
 Eccles Institute of Human Genetics, University of Utah

15 N 2030 EAST RM 2100, Salt Lake City, UT 84112-5330, USA
Tel: 801 581-4422
Fax: 801 581-7796
Email: alersou@genetics.utah.edu
Seq primer: SP6.

```
seq primer: 5'G.
FEATURES
            Location/Qualifiers
            source             1..419
```

```

/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon.7227"
/dev_stage="embryonic 8-12 hr post-fertilization"
/clone_lib="Drosophila 8-12 hr embryonic cDNA library"
/notes="vector: plasmid pNB; see Brown, N.H. and Kafatos, F.C. (1988) Functional cDNA libraries from Drosophila embryos. J. Mol Biol 203, 425-37."

```

ORIGIN	
Query Match	84.0%; Score 16.8; DB 13; Length 419;
Best Local Similarity	90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative	0; Mismatches 2; Indels 0; Gaps 0;

1 GCGCGGCACCACTTCGACTT 20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
8

RESULT 25
 BI163420
 LOCUS
 DEFINITION
 BI163420 427 bp mRNA linear EST 09-JUL-2001
 REO2821.5prime RE *Drosophila* melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone REO2821.5 similar to Guf:
 Fnan0016747 'enzyme inhibitor', located on: 2R 48F1-48F1.1;
 04/11/2001, mRNA sequence.

ACCESSION	BT163420
VERSION	BT163420.1
KEYWORDS	GI:14629235
EST.	
SOURCE	Drosophila melanogaster
ORGANISM	Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 427)

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Mira, S., Mungall, C. J., Nuncio, J., Pacleb, J., Parasas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin, G. M.

TITLE	BDGP/HIMI RE Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP

Bruce
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AEO03823: arm:2R [6997351,7260081]
 estimated-cyto:48DB-48F5: 04/11/2001
 Plate: RE.28 row: B column: 9
 High quality sequence stop: 425.

```

FEATURES
source
Location/Qualifiers
1. .427
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE02821"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DHS-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pfic-1"
/notes="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Pier Carninci at

```

the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

Query Match	84.0%;	Score 16.8;	DB 12;	Length 427;
Best Local Similarity	90.0%;	Pred. No. 1.6e+04;		
Matches 18:	Conservative	0:	Mismatches 2:	Indels 0:
				Gaps 0:

Qy 1 GCGCGCACCTCGACTT 20
 | | | | |
Db 237 GCTCGCACCATCTCGACTT 256

20 **21**

RESULT 26
BI213953

LOCUS	BI213953	427 bp	mRNA	linear	EST 11-JUL-2001
DEFINITION	<p>RE19610.5prime RE 'Drosophila melanogaster, normalized Embryo pFLc-1 Drosophila melanogaster cDNA clone RE19610.5 similar to gnf; FBAN0016747 'enzyme inhibitor', located on: 2R 48F1-48F1; 04/12/2001 mRNA sequence.</p>				

ACCESSION	E1213953	
VERSION	E1213953.1	GI:14691677
KEYWORDS	EST.	
SOURCE	Drosophila melanogaster	
ORGANISM	Drosophila melanogaster	(fruit fly)

REFERENCES

1. (bases 1 to 427)

Ephyrdoidea; Drosophilidae; Drosophila.

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

1. Classes 1 to 12, 13 to 17, 18 to 22, 23 to 27, 28 to 32, 33 to 37, 38 to 42, 43 to 47, 48 to 52, 53 to 57, 58 to 62, 63 to 67, 68 to 72, 73 to 77, 78 to 82, 83 to 87, 88 to 92, 93 to 97, 98 to 102, 103 to 107, 108 to 112, 113 to 117, 118 to 122, 123 to 127, 128 to 132, 133 to 137, 138 to 142, 143 to 147, 148 to 152, 153 to 157, 158 to 162, 163 to 167, 168 to 172, 173 to 177, 178 to 182, 183 to 187, 188 to 192, 193 to 197, 198 to 202, 203 to 207, 208 to 212, 213 to 217, 218 to 222, 223 to 227, 228 to 232, 233 to 237, 238 to 242, 243 to 247, 248 to 252, 253 to 257, 258 to 262, 263 to 267, 268 to 272, 273 to 277, 278 to 282, 283 to 287, 288 to 292, 293 to 297, 298 to 302, 303 to 307, 308 to 312, 313 to 317, 318 to 322, 323 to 327, 328 to 332, 333 to 337, 338 to 342, 343 to 347, 348 to 352, 353 to 357, 358 to 362, 363 to 367, 368 to 372, 373 to 377, 378 to 382, 383 to 387, 388 to 392, 393 to 397, 398 to 402, 403 to 407, 408 to 412, 413 to 417, 418 to 422, 423 to 427, 428 to 432, 433 to 437, 438 to 442, 443 to 447, 448 to 452, 453 to 457, 458 to 462, 463 to 467, 468 to 472, 473 to 477, 478 to 482, 483 to 487, 488 to 492, 493 to 497, 498 to 502, 503 to 507, 508 to 512, 513 to 517, 518 to 522, 523 to 527, 528 to 532, 533 to 537, 538 to 542, 543 to 547, 548 to 552, 553 to 557, 558 to 562, 563 to 567, 568 to 572, 573 to 577, 578 to 582, 583 to 587, 588 to 592, 593 to 597, 598 to 602, 603 to 607, 608 to 612, 613 to 617, 618 to 622, 623 to 627, 628 to 632, 633 to 637, 638 to 642, 643 to 647, 648 to 652, 653 to 657, 658 to 662, 663 to 667, 668 to 672, 673 to 677, 678 to 682, 683 to 687, 688 to 692, 693 to 697, 698 to 702, 703 to 707, 708 to 712, 713 to 717, 718 to 722, 723 to 727, 728 to 732, 733 to 737, 738 to 742, 743 to 747, 748 to 752, 753 to 757, 758 to 762, 763 to 767, 768 to 772, 773 to 777, 778 to 782, 783 to 787, 788 to 792, 793 to 797, 798 to 802, 803 to 807, 808 to 812, 813 to 817, 818 to 822, 823 to 827, 828 to 832, 833 to 837, 838 to 842, 843 to 847, 848 to 852, 853 to 857, 858 to 862, 863 to 867, 868 to 872, 873 to 877, 878 to 882, 883 to 887, 888 to 892, 893 to 897, 898 to 902, 903 to 907, 908 to 912, 913 to 917, 918 to 922, 923 to 927, 928 to 932, 933 to 937, 938 to 942, 943 to 947, 948 to 952, 953 to 957, 958 to 962, 963 to 967, 968 to 972, 973 to 977, 978 to 982, 983 to 987, 988 to 992, 993 to 997, 998 to 1002, 1003 to 1007, 1008 to 1012, 1013 to 1017, 1018 to 1022, 1023 to 1027, 1028 to 1032, 1033 to 1037, 1038 to 1042, 1043 to 1047, 1048 to 1052, 1053 to 1057, 1058 to 1062, 1063 to 1067, 1068 to 1072, 1073 to 1077, 1078 to 1082, 1083 to 1087, 1088 to 1092, 1093 to 1097, 1098 to 1102, 1103 to 1107, 1108 to 1112, 1113 to 1117, 1118 to 1122, 1123 to 1127, 1128 to 1132, 1133 to 1137, 1138 to 1142, 1143 to 1147, 1148 to 1152, 1153 to 1157, 1158 to 1162, 1163 to 1167, 1168 to 1172, 1173 to 1177, 1178 to 1182, 1183 to 1187, 1188 to 1192, 1193 to 1197, 1198 to 1202, 1203 to 1207, 1208 to 1212, 1213 to 1217, 1218 to 1222, 1223 to 1227, 1228 to 1232, 1233 to 1237, 1238 to 1242, 1243 to 1247, 1248 to 1252, 1253 to 1257, 1258 to 1262, 1263 to 1267, 1268 to 1272, 1273 to 1277, 1278 to 1282, 1283 to 1287, 1288 to 1292, 1293 to 1297, 1298 to 1302, 1303 to 1307, 1308 to 1312, 1313 to 1317, 1318 to 1322, 1323 to 1327, 1328 to 1332, 1333 to 1337, 1338 to 1342, 1343 to 1347, 1348 to 1352, 1353 to 1357, 1358 to 1362, 1363 to 1367, 1368 to 1372, 1373 to 1377, 1378 to 1382, 1383 to 1387, 1388 to 1392, 1393 to 1397, 1398 to 1402, 1403 to 1407, 1408 to 1412, 1413 to 1417, 1418 to 1422, 1423 to 1427, 1428 to 1432, 1433 to 1437, 1438 to 1442, 1443 to 1447, 1448 to 1452, 1453 to 1457, 1458 to 1462, 1463 to 1467, 1468 to 1472, 1473 to 1477, 1478 to 1482, 1483 to 1487, 1488 to 1492, 1493 to 1497, 1498 to 1502, 1503 to 1507, 1508 to 1512, 1513 to 1517, 1518 to 1522, 1523 to 1527, 1528 to 1532, 1533 to 1537, 1538 to 1542, 1543 to 1547, 1548 to 1552, 1553 to 1557, 1558 to 1562, 1563 to 1567, 1568 to 1572, 1573 to 1577, 1578 to 1582, 1583 to 1587, 1588 to 1592, 1593 to 1597, 1598 to 1602, 1603 to 1607, 1608 to 1612, 1613 to 1617, 1618 to 1622, 1623 to 1627, 1628 to 1632, 1633 to 1637, 1638 to 1642, 1643 to 1647, 1648 to 1652, 1653 to 1657, 1658 to 1662, 1663 to 1667, 1668 to 1672, 1673 to 1677, 1678 to 1682, 1683 to 1687, 1688 to 1692, 1693 to 1697, 1698 to 1702, 1703 to 1707, 1708 to 1712, 1713 to 1717, 1718 to 1722, 1723 to 1727, 1728 to 1732, 1733 to 1737, 1738 to 1742, 1743 to 1747, 1748 to 1752, 1753 to 1757, 1758 to 1762, 1763 to 1767, 1768 to 1772, 1773 to 1777, 1778 to 1782, 1783 to 1787, 1788 to 1792, 1793 to 1797, 1798 to 1802, 1803 to 1807, 1808 to 1812, 1813 to 1817, 1818 to 1822, 1823 to 1827, 1828 to 1832, 1833 to 1837, 1838 to 1842, 1843 to 1847, 1848 to 1852, 1853 to 1857, 1858 to 1862, 1863 to 1867, 1868 to 1872, 1873 to 1877, 1878 to 1882, 1883 to 1887, 1888 to 1892, 1

TITLE	JOURNAL	COMMENT
BDGP/HMI RE Drosophila EST Project	Unpublished (2001)	Contact: Stapleton, M. BDGP

EDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003823: arm:2R 16997351,7260081]
estimated-cyto:48D8-48F5: 04/12/2001
Plate: RE.196 row: A column: 10
High quality sequence stop: 250.

```

FEATURES
source
1. .427
/organism="Drosophila melanogaster"
/mol_type="mRNA"
location/Qualifiers
high_quality sequence RefSeq: 230.

```

```

/abn_cytoB="mRNA"
/abn_xref="taxon: 7227"
/clone="RE19610"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFLC-1"
/note="Organ: embryo; Vector: pFLC1; Site:1:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

```

ORIGIN

Query Match	84.0%;	Score 16.8;	DB 12;	Length 427;
Best Local Similarity	90.0%;	Pred. No. 1.6e+04;		
Matches 18;	Conservative	0;	Mismatches 2;	Indels 0;
			Gaps	0;

CDC Recommendations. Accessed 02/04/19.

Qy 1 GGGCGCACCACTCGACTT 20
|||
Db 238 GCTGGCGACCACTCGACTT 257
|||

```

RESULT 27
BI375669
LOCUS
DEFINITION
  RE63532.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE63532 5 similar to gtf:
  FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
  located on: 2R 48F1-48F1; 05/16/2001, mRNA sequence.
ACCESSION
  BI375669
VERSION
  BI375669.1 GI:15071697
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, estofruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5: 05/16/2001
  Plate: RE 635 row: C column: 8
  High quality sequence stop: 426.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE63532"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 238 GCTCCGACCACTCGACTT 257

RESULT 28
BI484665
LOCUS
DEFINITION
  RE67806.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE67806 5 similar to gtf:
  FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
  located on: 2R 48F1-48F1; 05/22/2001, mRNA sequence.
ACCESSION
  BI484665
VERSION
  BI484665.1 GI:15324957
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, estofruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5: 05/16/2001
  Plate: RE 635 row: C column: 8
  High quality sequence stop: 426.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE63532"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 238 GCTCCGACCACTCGACTT 257

```

```

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, estofruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5: 05/22/2001
  Plate: RE 678 row: A column: 6
  High quality sequence stop: 379.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE67806"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 237 GCTCCGACCACTCGACTT 256

RESULT 29
BI365791
LOCUS
DEFINITION
  RE51171.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE51171 5 similar to gtf:
  FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
  05/14/2001, mRNA sequence.
ACCESSION
  BI365791
VERSION
  BI365791.1 GI:15061819
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 428)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP

```

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/14/2001
Plate: RE.511 row: F column: 11
High quality sequence stop: 426.

FEATURES
source

1. .428
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE51171"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 428;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGACTT 20
|||
Db 238 GCTCGGACCACTCTCGACTT 257

RESULT 30

BI366732

LOCUS

DEFINITION

BI366732 428 bp mRNA linear EST 01-AUG-2001
RE52331.1 Spine RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE52331 5 similar to gtf:
FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1.1;
05/14/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 428)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and
Rubin, G. M.

TITLE

JOURNAL

COMMENT

BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/14/2001
Plate: RE.523 row: C column: 7
High quality sequence stop: 426.
Location/Qualifiers

FEATURES
source

1. .428
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"

/clone="RE52331"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 428;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGACTT 20
|||
Db 238 GCTCGGACCACTCTCGACTT 257

Search completed: June 20, 2004, 14:13:52
Job time : 951.603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 22.1164 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-13
Perfect score: 20
Sequence: 1 ggcgcgcaccactcagctt 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	20	100.0	4403765	3 US-09-103-840A-2
2	20	100.0	4411529	3 US-09-103-840A-1
3	16.4	82.0	3600	4 US-09-657-042A-3
4	15.8	79.0	987	4 US-09-252-991A-7285
5	15.8	79.0	1083	4 US-09-489-039A-2785
6	15.8	79.0	1368	4 US-09-252-991A-7174
7	15.8	79.0	1575	4 US-09-252-991A-7330
8	15.4	77.0	114	4 US-09-536-977-37
9	15.4	77.0	272	4 US-09-536-977-59
10	15.4	77.0	453	4 US-09-252-991A-12478
11	15.4	77.0	495	4 US-09-252-991A-12406
12	15.4	77.0	525	4 US-09-252-991A-12111
13	15.4	77.0	690	4 US-09-252-991A-12360
14	15.4	77.0	924	4 US-09-252-991A-12250
15	15.4	77.0	1011	4 US-09-252-991A-12311
16	15.4	77.0	1038	4 US-09-536-977-75
17	15.4	77.0	1085	4 US-09-522-714-15
18	15.4	77.0	1143	4 US-09-252-991A-12185
19	15.4	77.0	1404	4 US-09-252-991A-13158
20	15.4	77.0	2163	4 US-09-252-991A-15435
21	15.4	77.0	2469	4 US-09-536-977-71
22	15.4	77.0	2844	4 US-09-252-991A-12788
23	15.4	77.0	3342	4 US-09-252-991A-12644
24	15.4	77.0	6972	4 US-09-595-684B-38
25	15.4	77.0	7011	4 US-09-268-153-9
26	15.4	8309	4	US-09-620-312D-1083
27	15.4	77.0	4403765	3 US-09-103-840A-2

C 28	15.4	77.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 29	15.2	76.0	333	4	US-09-621-976-1316	Sequence 1316, Ap
C 30	15.2	76.0	588	4	US-09-252-991A-8968	Sequence 8968, Ap
C 31	15.2	76.0	606	4	US-09-252-991A-9120	Sequence 9120, Ap
C 32	15.2	76.0	627	4	US-09-252-991A-9038	Sequence 9038, Ap
C 33	15.2	76.0	673	4	US-09-833-381-800	Sequence 800, App
C 34	15.2	76.0	945	4	US-09-489-039A-2349	Sequence 2349, Ap
C 35	15.2	76.0	1014	4	US-09-489-039A-6032	Sequence 6032, Ap
C 36	15.2	76.0	1029	4	US-09-252-991A-153	Sequence 153, App
C 37	15.2	76.0	1038	4	US-09-252-991A-13643	Sequence 13643, A
C 38	15.2	76.0	1110	4	US-09-252-991A-173	Sequence 173, App
C 39	15.2	76.0	1161	4	US-09-252-991A-9202	Sequence 9202, Ap
C 40	15.2	76.0	1209	4	US-09-252-991A-8857	Sequence 8857, Ap
C 41	15.2	76.0	1221	6	5212296-16	Patent No. 5212296
C 42	15.2	76.0	1251	4	US-09-252-991A-13795	Sequence 13795, A
C 43	15.2	76.0	1266	4	US-09-252-991A-183	Sequence 183, App
C 44	15.2	76.0	1611	4	US-09-252-991A-183	Patent No. 5212296
C 45	15.2	76.0	1879	6	5212296-5	Sequence 5, Appli
C 46	15.2	76.0	2932	3	US-08-999-774A-5	Sequence 165, App
C 47	15.2	76.0	2933	4	US-09-149-476-185	Sequence 296, App
C 48	15.2	76.0	3276	4	US-09-149-476-298	Sequence 298, Appl
C 49	15.2	76.0	3768	4	US-09-566-921-50	Sequence 50, Appl
C 50	15.2	76.0	7812	3	US-09-368-590-1	Sequence 1, Appli
C 51	15.2	76.0	536165	4	US-09-214-808-1	Sequence 1, Appli
C 52	14.8	74.0	30	3	US-08-817-787-30	Sequence 30, Appl
C 53	14.8	74.0	41	4	US-09-463-067-19	Sequence 19, Appl
C 54	14.8	74.0	182	3	US-08-818-112-29	Sequence 29, Appl
C 55	14.8	74.0	182	4	US-08-818-111-29	Sequence 29, Appl
C 56	14.8	74.0	182	4	US-09-056-556-29	Sequence 29, Appl
C 57	14.8	74.0	182	4	US-09-072-596-29	Sequence 29, Appl
C 58	14.8	74.0	182	4	US-09-072-967-29	Sequence 29, Appl
C 59	14.8	74.0	272	4	US-09-313-294A-2605	Sequence 2605, Ap
C 60	14.8	74.0	459	4	US-09-489-039A-3617	Sequence 3617, Ap
C 61	14.8	74.0	663	4	US-09-252-991A-1065	Sequence 1065, Ap
C 62	14.8	74.0	687	4	US-09-252-991A-11810	Sequence 11810, A
C 63	14.8	74.0	814	4	US-09-187-999-18	Sequence 18, Appl
C 64	14.8	74.0	870	4	US-09-252-991A-15989	Sequence 15989, A
C 65	14.8	74.0	1062	4	US-09-252-991A-16198	Sequence 16198, A
C 66	14.8	74.0	1212	4	US-09-489-039A-4706	Sequence 4706, Ap
C 67	14.8	74.0	1233	4	US-09-489-039A-4691	Sequence 4691, Ap
C 68	14.8	74.0	1301	2	US-08-641-314C-1	Sequence 1, Appli
C 69	14.8	74.0	1416	4	US-09-252-991A-1100	Sequence 1100, Ap
C 70	14.8	74.0	1425	4	US-09-252-991A-11513	Sequence 11513, A
C 71	14.8	74.0	1434	4	US-09-489-039A-3399	Sequence 3399, Ap
C 72	14.8	74.0	1458	4	US-09-489-039A-1373	Sequence 1373, Ap
C 73	14.8	74.0	1470	4	US-09-252-991A-896	Sequence 896, App
C 74	14.8	74.0	1511	2	US-08-809-763-3	Sequence 3, Appli
C 75	14.8	74.0	1511	3	US-08-956-253-3	Sequence 3, Appli
C 76	14.8	74.0	1512	4	US-09-252-991A-1138	Sequence 1138, Ap
C 77	14.8	74.0	1555	2	US-08-696-376-1	Sequence 1, Appli
C 78	14.8	74.0	4326	4	US-09-252-991A-11572	Sequence 11572, A
C 79	14.8	74.0	4473	4	US-09-252-991A-11871	Sequence 11871, A
C 80	14.8	74.0	33529	3	US-09-144-085-3	Sequence 3, Appli
C 81	14.8	74.0	36033	4	US-08-311-731A-124	Sequence 124, App
C 82	14.8	74.0	68750	3	US-09-335-409-1	Sequence 1, Appli
C 83	14.8	74.0	68750	4	US-09-568-102-1	Sequence 1, Appli
C 84	14.8	74.0	68750	4	US-09-567-969-1	Sequence 1, Appli
C 85	14.8	74.0	68750	4	US-09-568-480-1	Sequence 1, Appli
C 86	14.8	74.0	68750	4	US-09-568-486-1	Sequence 1, Appli
C 87	14.8	74.0	68750	4	US-09-568-472-1	Sequence 1, Appli
C 88	14.8	74.0	68750	4	US-09-567-899-1	Sequence 2, Appli
C 89	14.8	74.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
C 90	14.4	72.0	45	3	US-08-406-855A-9	Sequence 9, Appli
C 91	14.4	72.0	45	3	US-09-206-899-9	Sequence 9, Appli
C 92	14.4	72.0	45	3	US-09-688-415-14	Sequence 14, Appl
C 93	14.4	72.0	376	4	US-09-621-976-16394	Sequence 16394, A
C 94	14.4	72.0	399	4	US-09-489-039A-6019	Sequence 6019, Ap
C 95	14.4	72.0	406	4	US-09-621-976-16393	Sequence 16393, A
C 96	14.4	72.0	477	4	US-09-252-991A-10565	Sequence 10565, A
C 97	14.4	72.0	478	4	US-09-280-116-174	Sequence 174, App
C 98	14.4	72.0	483	4	US-09-621-976-2791	Sequence 2791, Ap
C 99	14.4	72.0	516	4	US-09-252-991A-16362	Sequence 16362, A
C 100	14.4	72.0	810	4	US-09-489-039A-1365	Sequence 1365, Ap

247 14.2 71.0 6749 3 US-08-796-101-35 Sequence 35, Appl
248 14.2 71.0 6749 3 US-08-796-101-38 Sequence 38, Appl
249 14.2 71.0 6749 3 US-09-085-273-71 Sequence 71, Appl
250 14.2 71.0 6749 3 US-09-916-963-71 Sequence 71, Appl
251 14.2 71.0 9208 4 US-09-068-506-1 Sequence 1, Appl
252 14.2 71.0 9758 4 US-09-392-812A-5 Sequence 5, Appl
253 14.2 71.0 14796 3 US-08-975-080-35 Sequence 35, Appl
254 14.2 71.0 14796 3 US-09-630-706-10 Sequence 10, Appl
255 14.2 71.0 14796 3 US-09-496-694B-3 Sequence 3, Appl
256 14.2 71.0 33382 4 US-08-311-731A-121 Sequence 121, Appl
257 14.2 71.0 43240 2 US-08-804-227C-1 Sequence 1, Appl
258 14 70.0 489 4 US-09-252-991A-6875 Sequence 6875, Ap
259 14 70.0 624 4 US-09-252-991A-13629 Sequence 13629, A
260 14 70.0 801 4 US-09-252-991A-13819 Sequence 13819, A
261 14 70.0 885 4 US-09-489-039A-5473 Sequence 5473, Ap
262 14 70.0 1041 4 US-09-543-681A-321 Sequence 321, Ap
263 14 70.0 99916 4 US-09-816-095-3 Sequence 3, Appl
264 13.8 69.0 267 4 US-09-313-294A-6252 Sequence 6252, Ap
265 13.8 69.0 274 4 US-09-313-294A-5136 Sequence 5136, Ap
266 13.8 69.0 276 4 US-09-313-294A-5375 Sequence 5375, Ap
267 13.8 69.0 294 4 US-09-313-294A-5319 Sequence 5319, Ap
268 13.8 69.0 324 4 US-09-252-991A-14145 Sequence 14145, A
269 13.8 69.0 351 4 US-08-252-991A-4464 Sequence 4464, Ap
270 13.8 69.0 360 4 US-09-252-991A-12618 Sequence 12618, A
271 13.8 69.0 386 4 US-09-313-294A-4261 Sequence 4261, Ap
272 13.8 69.0 404 4 US-09-612-473-43 Sequence 43, Appl
273 13.8 69.0 435 4 US-09-252-991A-6670 Sequence 6670, Ap
274 13.8 69.0 456 4 US-09-252-991A-8984 Sequence 8984, Ap
275 13.8 69.0 473 4 US-08-621-976-14118 Sequence 14118, A
276 13.8 69.0 481 4 US-09-644-507B-5 Sequence 5, Appl
277 13.8 69.0 513 3 US-09-517-347-3 Sequence 3, Appl
278 13.8 69.0 513 3 US-09-517-347-7 Sequence 7, Appl
279 13.8 69.0 513 4 US-09-518-232A-3 Sequence 3, Appl
280 13.8 69.0 513 4 US-09-518-232A-7 Sequence 7, Appl
281 13.8 69.0 513 4 US-09-517-358A-3 Sequence 3, Appl
282 13.8 69.0 513 4 US-09-517-358A-7 Sequence 7, Appl
283 13.8 69.0 533 3 US-09-186-250-3 Sequence 3, Appl
284 13.8 69.0 533 3 US-09-186-250-7 Sequence 7, Appl
285 13.8 69.0 612 4 US-09-252-991A-2622 Sequence 2622, Ap
286 13.8 69.0 624 4 US-09-252-991A-14768 Sequence 14768, A
287 13.8 69.0 627 4 US-09-134-001C-36 Sequence 36, Appl
288 13.8 69.0 642 3 US-09-186-250-1 Sequence 1, Appl
289 13.8 69.0 642 3 US-09-517-347-1 Sequence 1, Appl
290 13.8 69.0 642 3 US-09-517-347-5 Sequence 5, Appl
291 13.8 69.0 642 4 US-09-518-232A-1 Sequence 1, Appl
292 13.8 69.0 642 4 US-09-518-232A-5 Sequence 5, Appl
293 13.8 69.0 642 4 US-09-517-358A-1 Sequence 1, Appl
294 13.8 69.0 642 4 US-09-517-358A-5 Sequence 5, Appl
295 13.8 69.0 652 3 US-09-186-250-5 Sequence 5, Appl
296 13.8 69.0 672 4 US-09-252-991A-13019 Sequence 13019, A
297 13.8 69.0 684 3 US-08-998-416-754 Sequence 754, App
298 13.8 69.0 696 4 US-09-408-020-17 Sequence 17, Appl
299 13.8 69.0 703 3 US-09-313-300-6 Sequence 6, Appl
300 13.8 69.0 738 4 US-09-252-991A-13548 Sequence 13548, A

ALIGNMENTS

RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2 4403765
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 20; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCACCTCGACTT 20
DB 3233708 GCGCGCACCACCTCGACTT 3233727

RESULT 2

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 20; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCACCTCGACTT 20
DB 3239432 GCGCGCACCACCTCGACTT 3239451

RESULT 3

US-09-657-042A-3
; Sequence 3, Application US/09657042A
; Patent No. 6329203
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
; FILE REFERENCE: RTS-0148
; CURRENT APPLICATION NUMBER: US/09/657,042A
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79) ... (3399)
US-09-657-042A-3


```

; SEQ ID NO 12111
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12111
Query Match      77.0%; Score 15.4;
Best Local Similarity 94.1%; Pred. No. 2

```

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
|||||
DB 521 CGCGCGCCACCTCGAC 505

RESULT 13

US-09-252-991A-12360
; Sequence 12360, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12360

; LENGTH: 690

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12360

Query Match

Best Local Similarity 77.0%; Score 15.4; DB 4; Length 690;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
|||||
DB 406 CGCGCGCCACCTCGAC 422

RESULT 14

US-09-252-991A-12250/c

; Sequence 12250, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12250

; LENGTH: 924

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12250

Query Match

Best Local Similarity 77.0%; Score 15.4; DB 4; Length 924;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
|||||
DB 534 CGCGCGCCACCTCGAC 518

RESULT 15

US-09-252-991A-12311

; Sequence 12311, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12311

; LENGTH: 1011

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12311

Query Match

Best Local Similarity 77.0%; Score 15.4; DB 4; Length 1011;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
|||||
DB 20 CGCGCGCCACCTCGAC 36

RESULT 16

US-09-536-977-75/c

; Sequence 75, Application US/09536977

; Patent No. 6649409

; GENERAL INFORMATION:

; APPLICANT: FOMSGAARD, ANDERS

; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT

; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED

; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE

; TITLE OF INVENTION: BX08 CONSTRUCTS

; FILE REFERENCE: 030307/0169

; CURRENT APPLICATION NUMBER: US/09/536,977

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/128,558

; PRIOR FILING DATE: 1999-04-09

; PRIOR APPLICATION NUMBER: DK PA1999 00427

; PRIOR FILING DATE: 1999-03-29

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 75

; LENGTH: 1038

; TYPE: DNA

; ORGANISM: Human immunodeficiency virus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1035)

US-09-536-977-75

Query Match

Best Local Similarity 77.0%; Score 15.4; DB 4; Length 1038;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCGCGCCACCTCGA 17
|||||
DB 972 CGCGCGCCACCTCGA 956

RESULT 17

US-09-522-714-15

; Sequence 15, Application US/09522714

; Patent No. 6563020

; GENERAL INFORMATION:

; APPLICANT: Simmons, Carl R.

; APPLICANT: Yalpani, Nasser

; TITLE OF INVENTION: Maize Chitinases and Their Use in

; TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants

```
; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522,714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125,915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1085
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)...(966)
US-09-522-714-15

Query Match      77.0%; Score 15.4; DB 4; Length 1085;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    ||||| ||||| ||||| |||||
Db 548 CGCGCGCACCACTCGAC 564

RESULT 18
US-09-252-991A-12185/c
; Sequence 12185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12185
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12185

Query Match      77.0%; Score 15.4; DB 4; Length 1143;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    ||||| ||||| ||||| |||||
Db 43 CGCGCGCACCACTCGAC 27

RESULT 19
US-09-252-991A-13158
; Sequence 13158, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13158

; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522,714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125,915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1085
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)...(966)
US-09-522-714-15

Query Match      77.0%; Score 15.4; DB 4; Length 1085;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    ||||| ||||| ||||| |||||
Db 548 CGCGCGCACCACTCGAC 564

RESULT 18
US-09-252-991A-12185/c
; Sequence 12185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12185
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12185

Query Match      77.0%; Score 15.4; DB 4; Length 1143;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    ||||| ||||| ||||| |||||
Db 43 CGCGCGCACCACTCGAC 27

RESULT 19
US-09-252-991A-13158
; Sequence 13158, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13158

; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522,714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125,915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1085
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)...(966)
US-09-522-714-15

Query Match      77.0%; Score 15.4; DB 4; Length 1404;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGA 17
    ||||| ||||| ||||| |||||
Db 679 GCGCTGCACCACTCGA 695

RESULT 20
US-09-252-991A-15435
; Sequence 15435, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15435
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1464)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15435

Query Match      77.0%; Score 15.4; DB 4; Length 2163;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    ||||| ||||| ||||| |||||
Db 1175 CGCGCGCACCACTCGAC 1191

RESULT 21
US-09-536-977-71/c
; Sequence 71, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FONGGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: (1)..(2450)
US-09-536-977-71

Query Match      77.0%; Score 15.4; DB 4; Length 2469;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGCCGACCACTCGA 17
||| |||||||||
Db 2397 GGCGTGCACCACCTCGA 2381

RESULT 22
US-09-252-991A-12788/c
; Sequence 12788, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12788
; LENGTH: 2844
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12788

Query Match      77.0%; Score 15.4; DB 4; Length 2844;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGCCGACCACTCGA 17
||| |||||||||
Db 2184 GGCGTGCACCACCTCGA 2168

RESULT 23
US-09-252-991A-12644/c
; Sequence 12644, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12644
; LENGTH: 3342
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12644

Query Match      77.0%; Score 15.4; DB 4; Length 3342;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGCCGACCACTCGA 17
||| |||||||||
Db 1690 GGCGTGCACCACCTCGA 1674

; NAME/KEY: CDS
; LOCATION: (1)..(2450)
US-09-595-684B-38/c
; Sequence 38, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugenii
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 6972
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-38

Query Match      77.0%; Score 15.4; DB 4; Length 6972;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGCGCACCACTCGA 17
||| |||||||||
Db 4866 GGCGCGCACCACTCGA 4850

RESULT 25
US-09-268-163-9
; Sequence 9, Application US/09268163B
; Patent No. 6353091
; GENERAL INFORMATION:
; APPLICANT: Lipscombe, Diane
; APPLICANT: Schorge, Stephanie
; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
; FILE REFERENCE: B1055/7000
; CURRENT APPLICATION NUMBER: US/09/268,163B
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 60/077,901
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 7011
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7008
US-09-268-163-9

Query Match      77.0%; Score 15.4; DB 4; Length 7011;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGCGCACCACTCGA 17
||| |||||||||
Db 5164 GGCGCGCACCACTCGA 5180

RESULT 26
US-09-620-312D-1083/c
; Sequence 1083, Application US/09620312D

```

Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma Yundong
APPLICANT: Wang, Duxui
APPLICANT: Wang, Zhwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_Fl_genes Version 1.0
SEQ ID NO 1083
LENGTH: 8309
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (162)..(2843)
US-09-620-312D-1083

Query Match 77.0%; Score 15.4; DB 4; Length 8309;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCGCGGCACCACTCGA 17
Db 4967 GCGCGGCACCACTCGA 4951

RESULT 27

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 77.0%; Score 15.4; DB 3; Length 4403765;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGGCACCACTCGAC 18
Db 1272177 CGCGGCACCACTCGAC 1272161

RESULT 28

US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 77.0%; Score 15.4; DB 3; Length 4411529;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGGCACCACTCGAC 18
Db 1272707 CGCGGCACCACTCGAC 1272691

RESULT 29

US-09-621-976-1316/c
Sequence 1316, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 1316
LENGTH: 333
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 13..333
US-09-621-976-1316

Query Match 76.0%; Score 15.2; DB 4; Length 333;
Best Local Similarity 85.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCGCGGCACCACTCGACTT 20
Db 285 GAGCGGCACCACTCGGCTT 266

RESULT 30
US-09-252-991A-8968
; Sequence 8968, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8968
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8968

Query Match 76.0%; Score 15.2; DB 4; Length 588;
Best Local Similarity 85.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGGCGGACACCTCGACTT 20
Db 102 GAGCGGACACCTCGCGTT 121

Search completed: June 20, 2004, 11:46:10
Job time : 37.2414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 113.122 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 gcgcgcaccaccctcgactt 20

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
C 1	20	100.0	867	13 US-10-282-122A-26184, A
C 2	20	100.0	870	13 US-10-282-122A-28622
C 3	18.4	92.0	849	13 US-10-282-122A-26037
C 4	18.4	92.0	849	13 US-10-282-122A-27722
C 5	17	85.0	495	13 US-10-282-122A-31936
C 6	16.8	84.0	846	15 US-10-260-238-1112
C 7	16.8	84.0	858	9 US-09-738-626-2272
C 8	16.8	84.0	1654	14 US-10-108-605-30
C 9	16.8	84.0	1913	13 US-10-424-599-94765
C 10	16.8	84.0	2163	14 US-10-108-605-32
C 11	16.8	84.0	3309400	9 US-09-738-626-1
C 12	16.4	82.0	965	16 US-10-260-238-946
C 13	16.4	82.0	1272	15 US-10-156-761-4971
C 14	16.4	82.0	1274	13 US-10-425-114-30185

Sequence 32442, A
Sequence 29, Appl
Sequence 3, Appl
Sequence 1503, Ap
Sequence 1504, Ap
Sequence 1, Appl
Sequence 2581, Ap
Sequence 4008, Ap
Sequence 32590, A
Sequence 29518, A
Sequence 1, Appl
Sequence 6097, Ap
Sequence 4, Appl
Sequence 23110, A
Sequence 13600, A
Sequence 26722, A
Sequence 1718, Ap
Sequence 641, Ap
Sequence 58, Appl
Sequence 1056, Ap
Sequence 356, Ap
Sequence 50, Appl
Sequence 363, Ap
Sequence 175, Ap
Sequence 47, Appl
Sequence 1895, Ap
Sequence 197, Ap
Sequence 13272, A
Sequence 324, Ap
Sequence 324, Ap
Sequence 4423, A
Sequence 55710, A
Sequence 27864, A
Sequence 12856, A
Sequence 1, Appl
Sequence 9, Appl
Sequence 5708, A
Sequence 7167, Ap
Sequence 15, Appl
Sequence 32575, A
Sequence 27498, A
Sequence 4039, Ap
Sequence 7376, Ap
Sequence 6926, Ap
Sequence 14822, A
Sequence 1, Appl
Sequence 9, Appl
Sequence 455, Ap
Sequence 281, Ap
Sequence 335, Ap
Sequence 1083, Ap
Sequence 1083, Ap
Sequence 348, Ap
Sequence 1, Appl
Sequence 645, Ap
Sequence 454, Ap
Sequence 42695, A
Sequence 179, Ap
Sequence 3402, Ap
Sequence 14229, A
Sequence 15225, A
Sequence 800, Ap
Sequence 71552, A
Sequence 127268, A
Sequence 6370, Ap
Sequence 127268, A
Sequence 82, Appl
Sequence 3993, Ap
Sequence 14184, A
Sequence 30193, A
Sequence 36, Appl
Sequence 39, Appl
Sequence 6526, Ap

16.4 82.0 1542 16 US-10-369-493-32442
16.4 82.0 3587 15 US-10-157-031-29
16.4 82.0 3600 13 US-10-380-126-3
16.4 82.0 8522 9 US-09-764-868-1503
16.4 82.0 14866 9 US-09-764-868-1504
16.4 82.0 9025608 15 US-10-156-761-1
16.4 80.0 293 9 US-09-294-09338-3581
16.4 80.0 1197 13 US-10-425-114-4008
16.4 80.0 2043 13 US-10-425-114-32590
16.4 80.0 446 10 US-09-918-995-29518
16.4 79.0 894 15 US-10-138-970A-1
16.4 79.0 1401 15 US-10-156-761-6097
16.4 79.0 1641 15 US-10-253-971-4
16.4 79.0 1747 13 US-10-425-114-23110
16.4 79.0 1776 13 US-10-425-114-13600
16.4 79.0 1781 13 US-10-425-114-26722
16.4 79.0 2250 16 US-10-104-047-1718
16.4 79.0 2706 16 US-10-094-749-641
16.4 79.0 2837 16 US-09-822-846-58
16.4 79.0 2859 16 US-10-108-260A-1056
16.4 79.0 8144 13 US-10-221-714A-356
16.4 79.0 10519 10 US-09-738-630-50
16.4 79.0 11523 13 US-10-221-714A-363
16.4 79.0 11523 15 US-10-239-676-175
16.4 79.0 11523 15 US-10-172-086-47
16.4 79.0 11523 15 US-10-311-455-1895
16.4 79.0 11523 15 US-10-240-453-197
16.4 79.0 343 13 US-10-424-599-13272
16.4 77.0 360 9 US-09-938-842A-324
16.4 77.0 360 11 US-09-938-842A-324
16.4 77.0 777 16 US-10-369-493-44423
16.4 77.0 779 13 US-10-424-599-55710
16.4 77.0 855 16 US-10-369-493-27964
16.4 77.0 861 13 US-10-282-122A-12856
16.4 77.0 939 12 US-10-362-327-1
16.4 77.0 939 12 US-10-362-327-9
16.4 77.0 941 13 US-10-424-599-55708
16.4 77.0 1005 15 US-10-156-761-7167
16.4 77.0 1085 15 US-10-304-928-15
16.4 77.0 1294 13 US-10-425-114-32975
16.4 77.0 1344 13 US-10-282-122A-27498
16.4 77.0 1401 9 US-09-815-242-4039
16.4 77.0 1401 13 US-10-282-122A-7376
16.4 77.0 1521 15 US-10-156-761-6926
16.4 77.0 1588 13 US-10-282-122A-14822
16.4 77.0 6431 12 US-10-437-421-1
16.4 77.0 7011 14 US-10-033-026-9
16.4 77.0 7049 13 US-10-087-192-455
16.4 77.0 7420 16 US-10-295-027-281
16.4 77.0 7863 15 US-10-084-817-335
16.4 77.0 8309 16 US-10-037-270-1083
16.4 77.0 8309 16 US-10-117-722-1083
16.4 77.0 8829 15 US-10-084-817-348
16.4 77.0 63158 15 US-10-292-198-1
16.4 77.0 82993 15 US-10-080-170-645
16.4 77.0 439892 13 US-10-087-192-454
16.4 76.0 194 13 US-10-424-599-42695
16.4 76.0 309 9 US-09-867-550-179
16.4 76.0 254 9 US-09-983-965-3402
16.4 76.0 597 13 US-10-425-114-14229
16.4 76.0 654 13 US-10-282-122A-15225
16.4 76.0 673 9 US-09-833-381-800
16.4 76.0 684 13 US-10-424-599-71552
16.4 76.0 714 13 US-10-027-632-127268
16.4 76.0 714 15 US-10-156-761-6370
16.4 76.0 714 16 US-10-027-632-127268
16.4 76.0 735 17 US-10-258-662-82
16.4 76.0 1002 16 US-10-260-238-3993
16.4 76.0 1054 13 US-10-425-114-14184
16.4 76.0 1086 13 US-10-282-122A-30193
16.4 76.0 1104 15 US-10-321-188-36
16.4 76.0 1105 15 US-10-321-188-39
16.4 76.0 1116 15 US-10-156-761-6526

C 88	15.2	76.0	1173	16	US-10-369-493-31555	Sequence 31555, A	14.8	74.0	1038	16	US-10-369-493-39798	Sequence 39798, A
C 89	15.2	76.0	1176	16	US-10-369-493-43218	Sequence 43218, A	14.8	74.0	1050	16	US-10-369-493-41301	Sequence 41301, A
C 90	15.2	76.0	1238	13	US-10-425-114-26641	Sequence 26641, A	14.8	74.0	1055	16	US-10-369-493-34346	Sequence 34346, A
C 91	15.2	76.0	1260	13	US-10-422-366-3	Sequence 3, Appli	14.8	74.0	1059	15	US-10-156-761-3160	Sequence 3160, Ap
C 92	15.2	76.0	1260	13	US-10-422-366-5	Sequence 5, Appli	14.8	74.0	1074	16	US-10-369-493-31921	Sequence 31921, A
C 93	15.2	76.0	1287	13	US-10-425-114-1563	Sequence 1563, Ap	14.8	74.0	1188	13	US-10-282-122A-31591	Sequence 31591, A
C 94	15.2	76.0	1306	16	US-10-369-493-43157	Sequence 43157, A	14.8	74.0	1254	15	US-10-214-446-25	Sequence 25, Appl
C 95	15.2	76.0	1317	16	US-10-028-445-4	Sequence 4, Appli	14.8	74.0	1380	13	US-10-417-700A-134	Sequence 134, App
C 96	15.2	76.0	1428	16	US-10-369-493-35927	Sequence 35927, A	14.8	74.0	1429	13	US-10-425-114-15471	Sequence 15471, A
C 97	15.2	76.0	1446	16	US-10-369-493-36028	Sequence 36028, A	14.8	74.0	1440	13	US-10-417-700A-70	Sequence 70, Appl
C 98	15.2	76.0	1548	13	US-10-156-761-5601	Sequence 5601, Ap	14.8	74.0	1489	13	US-10-425-114-14295	Sequence 14295, A
C 99	15.2	76.0	1548	13	US-10-156-761-5601	Sequence 5601, Ap	14.8	74.0	1489	13	US-10-425-114-14295	Sequence 14295, A
C 100	15.2	76.0	1722	15	US-10-156-761-4387	Sequence 4387, Ap	14.8	74.0	1556	15	US-10-081-872-121	Sequence 121, App
C 101	15.2	76.0	1826	13	US-10-028-445-1	Sequence 1, Appli	14.8	74.0	1556	15	US-10-385-305-121	Sequence 121, App
C 102	15.2	76.0	1959	13	US-10-282-122A-25712	Sequence 25712, A	14.8	74.0	1560	15	US-10-156-761-5073	Sequence 5073, Ap
C 103	15.2	76.0	1960	13	US-10-425-114-14258	Sequence 14258, A	14.8	74.0	1560	15	US-10-156-761-5073	Sequence 5073, Ap
C 104	15.2	76.0	2058	16	US-10-369-493-43117	Sequence 43117, A	14.8	74.0	1622	16	US-10-260-238-1073	Sequence 1073, Ap
C 105	15.2	76.0	2244	16	US-10-320-797-2288	Sequence 2288, Ap	14.8	74.0	1665	16	US-10-369-493-44323	Sequence 44323, A
C 106	15.2	76.0	2244	16	US-10-320-797-2288	Sequence 2288, Ap	14.8	74.0	1701	13	US-10-282-122A-32068	Sequence 32068, A
C 107	15.2	76.0	2289	13	US-10-282-122A-39166	Sequence 39166, Ap	14.8	74.0	1714	13	US-10-425-114-1753	Sequence 1753, Ap
C 108	15.2	76.0	2849	16	US-10-320-797-1288	Sequence 1288, Ap	14.8	74.0	1731	13	US-10-282-122A-26348	Sequence 26348, A
C 109	15.2	76.0	2894	13	US-10-282-122A-33613	Sequence 33613, A	14.8	74.0	1734	13	US-10-282-122A-28433	Sequence 28433, A
C 110	15.2	76.0	2933	10	US-09-809-391-165	Sequence 165, App	14.8	74.0	1790	13	US-10-425-114-3920	Sequence 3920, Ap
C 111	15.2	76.0	2933	13	US-09-882-171-165	Sequence 165, App	14.8	74.0	1857	13	US-10-260-238-1672	Sequence 1672, Ap
C 112	15.2	76.0	3195	9	US-09-922-217-1114	Sequence 1114, Ap	14.8	74.0	2000	16	US-10-260-238-2619	Sequence 2619, Ap
C 113	15.2	76.0	3195	14	US-10-025-380-1114	Sequence 1114, Ap	14.8	74.0	2067	15	US-10-156-761-6391	Sequence 6391, Ap
C 114	15.2	76.0	3198	16	US-10-314-434-98	Sequence 98, Appl	14.8	74.0	2148	15	US-10-156-761-3210	Sequence 3210, Ap
C 115	15.2	76.0	3276	10	US-09-809-391-298	Sequence 298, App	14.8	74.0	2162	13	US-10-425-114-25098	Sequence 25098, A
C 116	15.2	76.0	3276	10	US-09-882-171-298	Sequence 298, App	14.8	74.0	2162	13	US-10-425-114-25098	Sequence 25098, A
C 117	15.2	76.0	3276	13	US-10-164-049-639	Sequence 639, App	14.8	74.0	2247	13	US-09-764-891-7546	Sequence 7546, Ap
C 118	15.2	76.0	3776	16	US-10-264-049-639	Sequence 639, App	14.8	74.0	2248	13	US-10-282-122A-25458	Sequence 25458, A
C 119	15.2	76.0	4649	9	US-10-320-797-2288	Sequence 2288, App	14.8	74.0	2248	13	US-10-425-114-1870	Sequence 1870, A
C 120	15.2	76.0	8391	16	US-09-870-759-113	Sequence 113, App	14.8	74.0	2424	15	US-10-156-761-3443	Sequence 3443, Ap
C 121	15.2	76.0	8391	10	US-09-751-708A-113	Sequence 113, App	14.8	74.0	2463	13	US-10-282-122A-13810	Sequence 13810, A
C 122	15.2	76.0	9098	13	US-10-221-278-34	Sequence 34, Appl	14.8	74.0	2466	16	US-10-369-493-44476	Sequence 44476, A
C 123	15.2	76.0	9098	15	US-10-119-428-37	Sequence 37, Appl	14.8	74.0	2466	16	US-10-369-493-44476	Sequence 44476, A
C 124	15.2	76.0	9098	16	US-10-291-172-34	Sequence 34, Appl	14.8	74.0	2511	13	US-10-282-122A-11853	Sequence 11853, A
C 125	15.2	76.0	9210	9	US-09-712-363-100	Sequence 100, App	14.8	74.0	2511	13	US-10-282-122A-27674	Sequence 27674, A
C 126	15.2	76.0	9590	13	US-10-282-122A-33351	Sequence 33351, A	14.8	74.0	2581	15	US-10-156-761-5779	Sequence 5779, Ap
C 127	15.2	76.0	30042	15	US-10-080-170-650	Sequence 650, App	14.8	74.0	2981	16	US-10-260-238-1052	Sequence 1052, Ap
C 128	15.2	76.0	135638	15	US-10-314-657-1	Sequence 1, Appli	14.8	74.0	3083	16	US-10-094-749-1096	Sequence 1096, Ap
C 129	15.2	76.0	135638	15	US-10-314-657-1	Sequence 1, Appli	14.8	74.0	3190	16	US-10-260-238-1050	Sequence 1050, Ap
C 130	15.2	76.0	53165	10	US-09-939-964-1	Sequence 1, Appli	14.8	74.0	3668	10	US-09-927-837-21	Sequence 21, Appl
C 131	15.2	76.0	9025608	13	US-10-156-761-1	Sequence 1, Appli	14.8	74.0	3882	16	US-10-084-846A-49	Sequence 49, Appl
C 132	15	75.0	630	13	US-10-282-122A-14409	Sequence 14409, A	14.8	74.0	5110	9	US-09-764-846-310	Sequence 310, App
C 133	15	75.0	1608	15	US-10-156-761-555	Sequence 555, App	14.8	74.0	5110	15	US-10-091-483-310	Sequence 310, App
C 134	15	75.0	1947	13	US-10-425-114-20740	Sequence 20740, A	14.8	74.0	5489	15	US-10-311-455-2094	Sequence 2094, Ap
C 135	14.8	74.0	182	15	US-10-131-002-29	Sequence 29, Appl	14.8	74.0	7155	15	US-10-329-079-14	Sequence 14, Appl
C 136	14.8	74.0	396	13	US-10-084-843-29	Sequence 29, Appl	14.8	74.0	9970	16	US-09-764-877-2277	Sequence 2277, Ap
C 137	14.8	74.0	495	13	US-10-282-122A-25978	Sequence 25978, A	14.8	74.0	20922	13	US-10-242-515-2277	Sequence 2277, Ap
C 138	14.8	74.0	523	13	US-10-282-122A-22993	Sequence 22993, A	14.8	74.0	25085	15	US-09-942-025-14	Sequence 14, Appl
C 139	14.8	74.0	535	13	US-10-425-114-5650	Sequence 5650, Ap	14.8	74.0	25085	15	US-10-132-134-35	Sequence 35, Appl
C 140	14.8	74.0	596	13	US-09-770-152-771	Sequence 771, App	14.8	74.0	37360	15	US-10-329-079-6	Sequence 6, Appli
C 141	14.8	74.0	596	13	US-10-425-114-22439	Sequence 22439, A	14.8	74.0	50843	15	US-10-132-134-25	Sequence 25, Appl
C 142	14.8	74.0	640	15	US-10-260-238-2784	Sequence 2784, Ap	14.8	74.0	59816	16	US-10-084-846A-1	Sequence 1, Appli
C 143	14.8	74.0	651	15	US-10-156-761-3835	Sequence 3835, Ap	14.8	74.0	59816	16	US-10-084-846A-2	Sequence 2, Appli
C 144	14.8	74.0	656	15	US-10-029-386-2082	Sequence 2082, A	14.8	74.0	63686	13	US-10-087-192-466	Sequence 466, App
C 145	14.8	74.0	677	13	US-10-425-114-3123	Sequence 31232, A	14.8	74.0	67311	13	US-10-092-025-1	Sequence 1, Appli
C 146	14.8	74.0	713	13	US-10-037-632-27651	Sequence 27651, A	14.8	74.0	6750	14	US-10-014-717-1	Sequence 1, Appli
C 147	14.8	74.0	713	16	US-10-037-632-27651	Sequence 27651, A	14.8	74.0	71989	13	US-09-727-889-2	Sequence 2, Appli
C 148	14.8	74.0	737	13	US-10-425-114-22763	Sequence 22763, A	14.8	74.0	207433	16	US-10-277-216-5	Sequence 5, Appli
C 149	14.8	74.0	752	13	US-10-027-632-135346	Sequence 135346, A	14.8	74.0	207433	17	US-10-126-022-5	Sequence 5, Appli
C 150	14.8	74.0	752	13	US-10-027-632-135347	Sequence 135347, A	14.8	74.0	45	15	US-10-238-129-14	Sequence 14, Appl
C 151	14.8	74.0	752	16	US-10-027-632-135346	Sequence 135346, A	14.8	74.0	45	15	US-10-238-667-14	Sequence 14, Appl
C 152	14.8	74.0	864	15	US-10-156-761-3578	Sequence 3578, Ap	14.8	74.0	135	13	US-10-085-783A-20971	Sequence 20971, A
C 153	14.8	74.0	894	13	US-10-282-122A-11787	Sequence 11787, A	14.8	74.0	135	16	US-10-242-535A-20971	Sequence 20971, A
C 154	14.8	74.0	939	13	US-10-282-122A-14724	Sequence 14724, A	14.8	74.0	165	16	US-10-085-783A-15623	Sequence 15623, A
C 155	14.8	74.0	968	16	US-10-369-493-44335	Sequence 44335, A	14.8	74.0	177	13	US-10-085-783A-10151	Sequence 10151, A
C 156	14.8	74.0	993	16	US-10-084-846A-67	Sequence 67, Appl	14.8	74.0	246	16	US-10-085-783A-44778	Sequence 44778, A
C 157	14.8	74.0	1008	15	US-10-156-761-1973	Sequence 1973, Ap	14.8	74.0	246	16	US-10-242-535A-1689	Sequence 1689, Ap
C 158	14.8	74.0	1035	16	US-10-369-493-39047	Sequence 39047, A	14.8	74.0	269	13	US-10-260-238-1050	Sequence 1050, Ap
C 159	14.8	74.0	1035	16	US-10-369-493-39047	Sequence 39047, A	14.8	74.0	269	13	US-10-085-783A-32925	Sequence 32925, A
C 160	14.8	74.0	1038	16	US-10-369-493-39414	Sequence 39414, A	14.8	74.0	269	16	US-10-242-535A-32925	Sequence 32925, A

C 234	14.4	72.0	271	13	US-10-085-783A-31525	Sequence 31525, A
C 235	14.4	72.0	271	16	US-10-242-535A-31525	Sequence 31525, A
C 236	14.4	72.0	307	13	US-09-852-386-2	Sequence 2, Appl
C 237	14.4	72.0	308	9	US-09-960-352-8755	Sequence 8755, Ap
C 238	14.4	72.0	326	9	US-09-960-352-9664	Sequence 9664, Ap
C 239	14.4	72.0	331	14	US-10-032-159A-24	Sequence 24, Appl
C 240	14.4	72.0	336	13	US-10-085-783A-3213	Sequence 3213, A
C 241	14.4	72.0	336	16	US-10-242-535A-3213	Sequence 3213, A
C 242	14.4	72.0	341	9	US-09-960-352-10028	Sequence 10028, A
C 243	14.4	72.0	358	16	US-10-264-049-1766	Sequence 1766, Ap
C 244	14.4	72.0	363	13	US-10-085-783A-23016	Sequence 23016, A
C 245	14.4	72.0	363	16	US-10-242-535A-23016	Sequence 23016, A
C 246	14.4	72.0	373	13	US-10-085-783A-16568	Sequence 16568, A
C 247	14.4	72.0	373	13	US-10-085-783A-45716	Sequence 45716, A
C 248	14.4	72.0	373	16	US-10-242-535A-45716	Sequence 45716, A
C 249	14.4	72.0	374	13	US-10-242-535A-52065	Sequence 52065, A
C 250	14.4	72.0	374	16	US-10-085-783A-52065	Sequence 52065, A
C 251	14.4	72.0	390	10	US-09-971-392A-234	Sequence 234, App
C 252	14.4	72.0	390	13	US-10-296-115-583	Sequence 583, App
C 253	14.4	72.0	405	13	US-10-424-599-24175	Sequence 24175, A
C 254	14.4	72.0	428	13	US-10-156-761-1554	Sequence 1554, Ap
C 255	14.4	72.0	453	16	US-10-264-049-657	Sequence 657, App
C 256	14.4	72.0	460	16	US-09-918-995-1033	Sequence 1033, Ap
C 257	14.4	72.0	472	10	US-10-369-493-40619	Sequence 40619, A
C 258	14.4	72.0	480	16	US-09-918-995-19184	Sequence 19184, A
C 259	14.4	72.0	485	10	US-09-918-995-19750	Sequence 19750, A
C 260	14.4	72.0	488	10	US-09-918-995-2212	Sequence 2212, Ap
C 261	14.4	72.0	508	10	US-09-918-995-350	Sequence 350, App
C 262	14.4	72.0	524	9	US-09-925-299-350	Sequence 350, App
C 263	14.4	72.0	524	10	US-09-925-299-350	Sequence 350, App
C 264	14.4	72.0	624	15	US-10-156-761-5927	Sequence 5927, Ap
C 265	14.4	72.0	630	13	US-10-425-114-24043	Sequence 24043, A
C 266	14.4	72.0	633	15	US-10-156-761-1655	Sequence 1655, Ap
C 267	14.4	72.0	670	15	US-10-132-350-29	Sequence 29, Appl
C 268	14.4	72.0	670	15	US-10-132-350-31	Sequence 31, Appl
C 269	14.4	72.0	678	13	US-10-424-599-13285	Sequence 13285, A
C 270	14.4	72.0	688	13	US-10-424-599-129203	Sequence 129203, A
C 271	14.4	72.0	815	16	US-10-074-978A-23	Sequence 23, Appl
C 272	14.4	72.0	856	13	US-10-027-632-161018	Sequence 161018, A
C 273	14.4	72.0	856	13	US-10-027-632-161019	Sequence 161019, A
C 274	14.4	72.0	856	16	US-10-027-632-161018	Sequence 161018, A
C 275	14.4	72.0	856	16	US-10-027-632-161019	Sequence 161019, A
C 276	14.4	72.0	862	16	US-10-260-238-1324	Sequence 1324, Ap
C 277	14.4	72.0	870	15	US-10-156-761-3118	Sequence 3118, Ap
C 278	14.4	72.0	946	13	US-10-425-114-33935	Sequence 33935, A
C 279	14.4	72.0	977	15	US-10-029-386-24193	Sequence 24193, A
C 280	14.4	72.0	1002	15	US-10-156-761-4347	Sequence 4347, Ap
C 281	14.4	72.0	1088	13	US-10-425-114-2306	Sequence 2306, Ap
C 282	14.4	72.0	1094	16	US-10-158-057-27	Sequence 27, Appl
C 283	14.4	72.0	1101	13	US-10-382-122A-25531	Sequence 25531, A
C 284	14.4	72.0	1113	13	US-10-336-472-73	Sequence 73, Appl
C 285	14.4	72.0	1113	13	US-10-336-472-77	Sequence 77, Appl
C 286	14.4	72.0	1113	13	US-10-336-472-83	Sequence 83, Appl
C 287	14.4	72.0	1116	15	US-10-156-761-5828	Sequence 5828, Ap
C 288	14.4	72.0	1149	13	US-10-412-698B-991	Sequence 991, App
C 289	14.4	72.0	1149	16	US-10-374-780A-487	Sequence 487, App
C 290	14.4	72.0	1197	15	US-10-260-715-11	Sequence 11, Appl
C 291	14.4	72.0	1200	15	US-10-146-733-3	Sequence 3, Appl
C 292	14.4	72.0	1200	15	US-10-146-733-6	Sequence 6, Appl
C 293	14.4	72.0	1216	9	US-09-746-491-21	Sequence 21, Appl
C 294	14.4	72.0	1216	13	US-10-336-472-75	Sequence 75, Appl
C 295	14.4	72.0	1233	13	US-10-382-122A-23596	Sequence 23596, A
C 296	14.4	72.0	1286	9	US-09-746-491-56	Sequence 56, Appl
C 297	14.4	72.0	1286	9	US-09-746-491-57	Sequence 57, Appl
C 298	14.4	72.0	1286	16	US-10-173-999-41	Sequence 41, Appl
C 299	14.4	72.0	1389	15	US-10-156-761-3920	Sequence 3920, Ap
C 300	14.4	72.0	1395	16	US-10-369-493-44291	Sequence 44291, A

ALIGNMENTS

RESULT 1

US-10-282-122A-26184/c
; Sequence 26184, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SEQ ID NO 26184
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26184

Query Match 100.0%; Score 20; DB 13; Length 867;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGCGCACCACTCGACTT 20
Db 36 GCGCGCACCACTCGACTT 17
|||||

RESULT 2

US-10-282-122A-28622/c
; Sequence 28622, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28622
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28622

Query Match          100.0%; Score 20; DB 13; Length 870;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACTT 20
DB 36 GCGCGGCACCACTCGACTT 17

RESULT 3
US-10-282-122A-26037/c
; Sequence 26037, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26037
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-26037

Query Match          92.0%; Score 18.4; DB 13; Length 849;
Best Local Similarity 95.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACTT 20
DB 36 GCGCGGCACCACTCGACTT 17

RESULT 4
US-10-282-122A-27722/c
; Sequence 27722, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27722
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-27722

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27722
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Mycobacterium leprae
US-10-282-122A-27722

Query Match          92.0%; Score 18.4; DB 13; Length 849;
Best Local Similarity 95.0%; Pred.No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  GCGCGGACACCACTCGACTT 20
          |||||
Db       36  GCGCGGACACCACTCTACTT 17

RESULT 5
US-10-282-122A-31936/c
; Sequence 31936, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31936
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-31936

Query Match          85.0%; Score 17; DB 13; Length 495;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3  GCGGACACCACTCGACT 19
          |||||
Db       277  GCGGACACCACTCGACT 261

```

TYPE: DNA
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-2272

Query Match 84.0%; Score 16.8; DB 9; Length 858;
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
 |||||
 DB 36 GGGCGGCACCACTCGACTT 17

RESULT 8

US-10-108-605-30
 ; Sequence 30, Application US/10108605
 ; Publication No. US20020160934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broadus, Julie
 ; APPLICANT: Stam, Lynn
 ; APPLICANT: Bachmann, Jane
 ; APPLICANT: Kamdar, Kim
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 ; FILE REFERENCE: 31133B
 ; CURRENT APPLICATION NUMBER: US/10/108,605
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US 09/761,142
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/176,418
 ; PRIOR FILING DATE: 2000-01-14
 ; NUMBER OF SEQ ID NOS: 361
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 30
 ; TYPE: DNA
 ; LENGTH: 1654
 ; ORGANISM: Drosophila melanogaster
 US-10-108-605-30

Query Match 84.0%; Score 16.8; DB 14; Length 1654;
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
 |||||
 DB 239 GCTCGGCACCACTCGACTT 258

RESULT 9

US-10-424-599-94765
 ; Sequence 94765, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kowalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 94765
 ; LENGTH: 1913
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_WRT3847_56584C.1
 US-10-424-599-94765

Query Match 84.0%; Score 16.8; DB 13; Length 1913;
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGGCGGCACCACTCGACTT 20
 |||||
 DB 172 GGGCGGCACCACTCGACTT 191

RESULT 10

US-10-108-605-32
 ; Sequence 32, Application US/10108605
 ; Publication No. US20020160934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broadus, Julie
 ; APPLICANT: Stam, Lynn
 ; APPLICANT: Bachmann, Jane
 ; APPLICANT: Kamdar, Kim
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 ; FILE REFERENCE: 31133B
 ; CURRENT APPLICATION NUMBER: US/10/108,605
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US 09/761,142
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/176,418
 ; PRIOR FILING DATE: 2000-01-14
 ; NUMBER OF SEQ ID NOS: 361
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 32
 ; TYPE: DNA
 ; LENGTH: 2163
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(2163)
 ; OTHER INFORMATION: n = a, t, c, or g
 US-10-108-605-32

Query Match 84.0%; Score 16.8; DB 14; Length 2163;
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20
 |||||
 DB 1082 GCTCGGCACCACTCGACTT 1101

RESULT 11

US-09-738-626-1
 ; Sequence 1, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIALI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0

```
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-1

Query Match      84.0%; Score 16.8; DB 9; Length 3309400;
Best Local Similarity 90.0%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGACTT 20
    ||| ||||| ||||| |||||
Db 2189135 GCGCGCGCACCACTCGACTT 2189154

RESULT 12
US-10-260-238-946/c
; Sequence 946, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Krepis, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 946
; LENGTH: 965
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-10-260-238-946

Query Match      82.0%; Score 16.4; DB 16; Length 965;
Best Local Similarity 94.4%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGAC 18
    ||| ||||| ||||| |||||
Db 36 GCGCGCGCACCACTCGTC 19

RESULT 13
US-10-156-761-4971
; Sequence 4971, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29

; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-1

Query Match      84.0%; Score 16.8; DB 9; Length 3309400;
Best Local Similarity 90.0%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGACTT 20
    ||| ||||| ||||| |||||
Db 2189135 GCGCGCGCACCACTCGACTT 2189154

RESULT 12
US-10-260-238-946/c
; Sequence 946, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Krepis, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 946
; LENGTH: 965
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-10-260-238-946

Query Match      82.0%; Score 16.4; DB 16; Length 965;
Best Local Similarity 94.4%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGAC 18
    ||| ||||| ||||| |||||
Db 36 GCGCGCGCACCACTCGTC 19

RESULT 13
US-10-156-761-4971
; Sequence 4971, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4971
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1272)
; US-10-156-761-4971

Query Match      82.0%; Score 16.4; DB 15; Length 1272;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGAC 18
    ||| ||||| ||||| |||||
Db 301 GCGCGCGCACCACTCGAC 318

RESULT 14
US-10-425-114-30185
; Sequence 30185, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30185
; LENGTH: 1274
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-OSFLM202096G05_FLI
; US-10-425-114-30185

Query Match      82.0%; Score 16.4; DB 13; Length 1274;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGAC 18
    ||| ||||| ||||| |||||
Db 818 GCGCGCGCACCACTCTAC 835

RESULT 15
US-10-369-493-32442/c
; Sequence 32442, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
```

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32442
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Ralstonia metallidurans
US-10-369-493-32442

Query Match 82.0%; Score 16.4; DB 16; Length 1542;
Best Local Similarity 94.4%; Pred. NO. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGAC 18
|||
Db 267 GCGCGCACCACTCGCC 250

RESULT 16

; Sequence 29, Application US/10157031
; Publication No. US2003010890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 3587
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-157-031-29

Query Match 82.0%; Score 16.4; DB 15; Length 3587;
Best Local Similarity 94.4%; Pred. NO. 2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
Db 3003 GCAGCACCACTCGACTT 3020

RESULT 17

US-10-380-126-3
; Sequence 3, Application US/10380126
; Publication No. US20040029824A1
; GENERAL INFORMATION:
; APPLICANT: Isis Pharmaceuticals, Inc.
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
; FILE REFERENCE: R1SP-0175
; CURRENT APPLICATION NUMBER: US/10/380,126
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 09/657,042
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79)...(3399)
US-10-380-126-3

Query Match 82.0%; Score 16.4; DB 13; Length 3600;

Best Local Similarity 94.4%; Pred. NO. 2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCGCACCACTCGACTT 20
|||
Db 3034 GCAGCACCACTCGACTT 3051

RESULT 18

US-09-764-868-1503/c
; Sequence 1503, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1503
; LENGTH: 8522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4173)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-1503

Query Match 82.0%; Score 16.4; DB 9; Length 8522;
Best Local Similarity 94.4%; Pred. NO. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
Db 7917 GCAGCACCACTCGACTT 7900

RESULT 19

US-09-764-868-1504/c
; Sequence 1504, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1504
; LENGTH: 14486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1504

Query Match 82.0%; Score 16.4; DB 9; Length 14486;
Best Local Similarity 94.4%; Pred. NO. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
Db 7921 GCAGCACCACTCGACTT 7904

RESULT 20

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIOYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 82.0%; Score 16.4; DB 15; Length 9025608;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCGCGGCACCACTCGAC 18
Db 6035826 GCGCGGCACCACTCGAC 6035843
RESULT 21
US-09-294-093B-2581/c
Sequence 2581, Application US/09294093B
Patent No. US20010051335A1
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath, V.
APPLICANT: Ito, Laura, Y.
APPLICANT: Sherman, Bradley, K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
FILE REFERENCE: PL-0009 US
CURRENT APPLICATION NUMBER: US/09/294,093B
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/082,567
PRIOR FILING DATE: April 21, 1998
NUMBER OF SEQ ID NOS: 6207
SOFTWARE: PERL Program
SEQ ID NO 2581
LENGTH: 293
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700346329H1
NAME/KEY: unsure
LOCATION: 2, 39, 109, 111, 115, 132
OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2581
Query Match 80.0%; Score 16; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACCACTCG 16
Db 91 GCGCGGCACCACTCG 76
RESULT 22

US-10-425-114-4008/c
Sequence 4008, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 4008
LENGTH: 1197
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700346329_FLI
US-10-425-114-4008
Query Match 80.0%; Score 16; DB 13; Length 1197;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACCACTCG 16
Db 90 GCGCGGCACCACTCG 75
RESULT 23
US-10-425-114-32590/c
Sequence 32590, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 32590
LENGTH: 2043
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: UC-ZMFLMO17007E01_FLI
US-10-425-114-32590
Query Match 80.0%; Score 16; DB 13; Length 2043;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACCACTCG 16
Db 1196 GCGCGGCACCACTCG 1181
RESULT 24
US-09-918-995-29518/c
Sequence 29518, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 29518

; LENGTH: 446

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc_feature

; LOCATION: (1)...(446)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-29518

Query Match 79.0%; Score 15.8; DB 10; Length 446;

Best Local Similarity 89.5%; Pred. No. 4.4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCGCGCACCTCGACTT 20

DB 114 GCGCGCACCTCGACTT 96

RESULT 25

US-10-138-970A-1

; Sequence 1, Application US/10138970A

; Publication No. US20030079255A1

; GENERAL INFORMATION:

; APPLICANT: Sariasiani, Sina

; APPLICANT: Tang, Xiao-Song

; APPLICANT: Qi, Wei Wei

; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

; FILE REFERENCE: CL1777

; CURRENT APPLICATION NUMBER: US/10/138,970A

; CURRENT FILING DATE: 2002-07-23

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 1

; LENGTH: 894

; TYPE: DNA

; ORGANISM: C. violaceum

US-10-138-970A-1

Query Match 79.0%; Score 15.8; DB 15; Length 894;

Best Local Similarity 89.5%; Pred. No. 4.2e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACT 19

DB 370 GAGCGGCACCTCGACT 388

RESULT 26

US-10-156-761-6097

; Sequence 6097, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 243-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 6097

; LENGTH: 1401

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1401)

US-10-156-761-6097

Query Match 79.0%; Score 15.8; DB 15; Length 1401;

Best Local Similarity 89.5%; Pred. No. 4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACT 19

DB 565 GCGCGCACCTCGACT 583

RESULT 27

US-10-253-971-4

; Sequence 4, Application US/10253971

; Publication No. US20030070192A1

; GENERAL INFORMATION:

; APPLICANT: GEORGES, FAWZY

; APPLICANT: DONG, JIN-ZHUO

; APPLICANT: KELLER, WILF

; APPLICANT: HUSSAIN, ATTA A. K.

; APPLICANT: SELVARAJ, GOPALAN

; APPLICANT: DATLA, RAJU

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT

; FILE REFERENCE: 73776-159D

; CURRENT APPLICATION NUMBER: US/10/253,971

; CURRENT FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: US'60/072156

; PRIOR FILING DATE: 1998-01-22

; PRIOR APPLICATION NUMBER: US 09/012453

; PRIOR FILING DATE: 1998-01-23

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1641

; TYPE: DNA

; ORGANISM: Arthrobacter pascens

US-10-253-971-4

Query Match 79.0%; Score 15.8; DB 15; Length 1641;

Best Local Similarity 89.5%; Pred. No. 4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACT 19

DB 436 GCGCGCACCTCGACT 454

RESULT 28

US-10-425-114-23110

; Sequence 23110, Application US/10425114

; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

Query Match 79.0%; Score 15.8; DB 13; Length 1781;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0;

Qy 1 GCGCGCACCACTCGACT 19
Db 325 GCGCGCACCTACCTCGAAT 343

Search completed: June 20, 2004, 17:45:46
Job time : 131.122 secs

RESULT 29
US-10-425-114-13600
; Sequence 13600, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 3841(53513)D
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128

```

/ SEQ ID NO 13600
/
/ LENGTH: 1776
/
/ TYPE: DNA
/
/ ORGANISM: Zea mays
/
/ FEATURE:
/
/ OTHER INFORMATION: Clone ID: LIB143-011-G1_FLI
US_-10-425-114-13600

```

Query Match	79.0%;	Score 15.8;	DB 13;	Length 1776;
Best Local Similarity	89.5%;	Pred. No. 4e+02;		
Matches 17: Conservative	0: Mismatches	2: Indels	0: Gaps	0: Gaps

Qy 1 GCGCGCACCCACCTCGACT 19
Db 286 GCGCGCACCTACCTCGAAT 304

RESULT 30
 US-10-425-114-26722
 ; Sequence 26722, Application US/10425114
 ; Publication No. US2004003488A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCES: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 597.376 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22
Sequence: 1 ccgggcaaaacctccgcctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_vrt.*
- 37: em_htg_wat.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	22	100.0	335	1	MBO298910	AJ298910 Mycobacte
C 2	22	100.0	335	1	MBO298911	AJ298911 Mycobacte
C 3	22	100.0	390	1	NMI1298907	AJ298907 Mycobacte
C 4	22	100.0	390	1	NMI1298909	AJ298909 Mycobacte
C 5	22	100.0	414	1	MBO298905	AJ298905 Mycobacte
C 6	22	100.0	469	1	NTU298905	AJ298906 Mycobacte
C 7	22	100.0	493	1	MBO298908	AJ298908 Mycobacte
C 8	22	100.0	10894	1	AE007121	AE007121 Mycobacte
C 9	22	100.0	318050	1	BX248344	BX248344 Mycobacte
C 10	22	100.0	348676	15	BX842581	BX842581 Mycobacte
C 11	18.8	85.5	242139	2	AC095431	AC095431 Rattus no
C 12	17.8	80.9	2629	3	AV119652	AV119652 Drosophil
C 13	17.8	80.9	46594	2	AC013897	AC013897 Drosophil
C 14	17.8	80.9	164415	3	AC023718	AC023718 Drosophil
C 15	17.8	80.9	166296	2	AC102929	AC102929 Mus muscu
C 16	17.8	80.9	229648	2	AC117358	AC117358 Rattus no
C 17	17.8	80.9	242397	2	AC097187	AC097187 Rattus no
C 18	17.8	80.9	254540	2	AC119383	AC119383 Rattus no
C 19	17.8	80.9	260335	2	AC115140	AC115140 Rattus no
C 20	17.8	80.9	286789	2	AC133723	AC133723 Rattus no
C 21	17.8	80.9	328128	3	AE003486	AE003486 Drosophil
C 22	17.4	79.1	163448	8	OSJN00127	AL606999 Oryza sat
C 23	17.4	79.1	176859	2	BX510331	BX510331 Danio rer
C 24	17.4	79.1	325612	2	AC095761	AC095761 Rattus no
C 25	17.2	78.2	2171	1	SEU43356	U43356 Salmonella
C 26	17.2	78.2	2171	1	SEU43357	U43357 Salmonella
C 27	17.2	78.2	2467	1	SEU43350	U43350 Salmonella
C 28	17.2	78.2	2467	1	SEU43351	U43351 Salmonella
C 29	17.2	78.2	2468	1	SEU43346	U43346 Salmonella
C 30	17.2	78.2	2468	1	SEU43347	U43347 Salmonella
C 31	17.2	78.2	5925	6	AX346478	AX346478 Sequence
C 32	17.2	78.2	64011	9	AC007642	AC007642 Homo sapi
C 33	17.2	78.2	65005	2	AC091277	AC091277 Mus muscu
C 34	17.2	78.2	77292	8	BX842596	BX842596 Neurospor
C 35	17.2	78.2	93508	10	AL606926	AL606926 Mouse DNA
C 36	17.2	78.2	189761	10	AC131177	AC131177 Mus muscu
C 37	17.2	78.2	192380	2	AC140346	AC140346 Mus muscu
C 38	17.2	78.2	204133	10	AC124504	AC124504 Mus muscu
C 39	17.2	78.2	217281	2	AC142259	AC142259 Mus muscu
C 40	17.2	78.2	217490	10	AC132402	AC132402 Mus muscu
C 41	17.2	78.2	225260	2	AC095002	AC095002 Rattus no
C 42	17.2	78.2	229299	2	AC097991	AC097991 Rattus no
C 43	17.2	78.2	262050	1	AP000985	AP000985 Sulfolobu
C 44	17.2	78.2	316869	2	AC124509	AC124509 Mus muscu
C 45	17.2	78.2	319431	2	AC097106	AC097106 Rattus no
C 46	17	77.3	46530	2	BX293554_5	Continuation (6 of
C 47	17	77.3	123556	9	HS071115	Z98950 Human DNA s
C 48	16.8	76.4	536	6	AX307280	AX307280 Sequence
C 49	16.8	76.4	6054	1	CANRDFGEN	Y09572 Corynebacte
C 50	16.8	76.4	77500	8	AP004166	AP004166 Oryza sat
C 51	16.8	76.4	84113	9	AC073063	AC073063 Homo sapi
C 52	16.8	76.4	156598	5	BX005055	BX005055 Zebrafish
C 53	16.8	76.4	185871	10	AC074311	AC074311 Mus muscu
C 54	16.8	76.4	219339	2	AC112623	AC112623 Rattus no
C 55	16.8	76.4	220280	2	AC145981	AC145981 Pan trogl
C 56	16.8	76.4	233152	2	AC134085	AC134085 Rattus no
C 57	16.8	76.4	237115	2	AC109127	AC109127 Rattus no
C 58	16.8	76.4	242845	2	AC126976	AC126976 Rattus no
C 59	16.8	76.4	246099	2	AC106365	AC106365 Rattus no
C 60	16.8	76.4	247570	2	AC130862	AC130862 Rattus no
C 61	16.8	76.4	259226	2	AC133656	AC133656 Rattus no
C 62	16.8	76.4	260199	2	AC110409	AC110409 Rattus no
C 63	16.8	76.4	279112	2	AC132048	AC132048 Rattus no
C 64	16.8	76.4	288925	2	AC146422	AC146422 Pan trogl
C 65	16.8	76.4	303846	2	AC103292	AC103292 Rattus no

66	16.8	76.4	340900	1	SME591791	AL591791 Sinorhizo	139	16.2	73.6	96482	2	AC136041	AC136041 Rattus no
67	16.4	74.5	771	6	AX677905 Sequence	AX677905 Sequence	C 140	16.2	73.6	97774	5	AL591399	AL591399 Zebraphish
68	16.4	74.5	2366	8	AK103682	AK103682 Oryza sat	C 141	16.2	73.6	110000	0	AC099174	AC099174 Rattus no
C 69	16.4	74.5	3000	1	R2TKILA	M62846 Plasmid RK2	C 142	16.2	73.6	110000	3	AE001572_1	Continuation (2 of
C 70	16.4	74.5	3086	1	R2KILA	M62426 Plasmid RK2	C 143	16.2	73.6	112732	9	AL160007	AL160007 Human DNA
71	16.4	74.5	3502	1	NGAJ2783	AJ002783 Neisseria	C 144	16.2	73.6	113164	2	AC136925	AC136925 Oryza sat
72	16.4	74.5	3738	9	BC005836	BC005836 Homo sapi	C 145	16.2	73.6	119014	2	AC142510	AC142510 Rattus no
73	16.4	74.5	3832	9	HSU12535	UI2535 Human epide	C 146	16.2	73.6	119909	2	AC068935	AC068935 Homo sapi
74	16.4	74.5	3887	9	BC030010	BC030010 Homo sapi	C 147	16.2	73.6	124921	2	AC133132	AC133132 Rattus no
C 75	16.4	74.5	10029	1	AE015044	AE015044 Shigella	C 148	16.2	73.6	130913	2	AC096869	AC096869 Rattus no
C 76	16.4	74.5	12106	1	AE001497	AE001497 Helicobac	C 149	16.2	73.6	138416	9	AC105218	AC105218 Homo sapi
C 77	16.4	74.5	12167	1	AE005184	AE005184 Escherich	C 150	16.2	73.6	141111	8	AP002481	AP002481 Oryza sat
78	16.4	74.5	14376	1	AE002531	AE002531 Neisseria	C 151	16.2	73.6	144094	9	AC087209	AC087209 Homo sapi
79	16.4	74.5	28043	2	AC014190	AC014190 Drosophil	C 152	16.2	73.6	145437	2	AC139985	AC139985 Rattus no
80	16.4	74.5	60099	1	BIACONGEN	L27758 Birmingham	C 153	16.2	73.6	156338	9	AC022887	AC022887 Homo sapi
C 81	16.4	74.5	64790	8	AP004863	AP004863 Oryza sat	C 154	16.2	73.6	158792	2	AC044874	AC044874 Homo sapi
C 82	16.4	74.5	75282	2	DMBR37M19	AL133495 Drosophil	C 155	16.2	73.6	165797	2	AC141664	AC141664 Apis mell
83	16.4	74.5	84353	9	AC092753	AC092753 Homo sapi	C 156	16.2	73.6	166015	9	AC113194	AC113194 Homo sapi
84	16.4	74.5	123518	8	AC123518	AC123518 Oryza sat	C 157	16.2	73.6	172937	10	AC121914	AC121914 Mus muscu
85	16.4	74.5	134011	9	AC005886	AC005886 b240g16,	C 158	16.2	73.6	177227	2	AC103580	AC103580 Trypanoso
86	16.4	74.5	141534	2	AP003685	AP003685 Oryza sat	C 159	16.2	73.6	181845	2	AC026586	AC026586 Homo sapi
87	16.4	74.5	162219	8	AC145321	AC145321 Oryza sat	C 160	16.2	73.6	182556	9	AC016879	AC016879 Homo sapi
88	16.4	74.5	163953	2	BX000689	BX000689 Danio rer	C 161	16.2	73.6	186107	2	AC023132	AC023132 Homo sapi
C 89	16.4	74.5	179009	9	AL355543	AL355543 Human DNA	C 162	16.2	73.6	186582	2	AC127620	AC127620 Rattus no
C 90	16.4	74.5	179892	3	AC023591	AC023591 Drosophil	C 163	16.2	73.6	187515	2	AC125383	AC125383 Rattus no
C 91	16.4	74.5	187223	2	AC008333	AC008333 Drosophil	C 164	16.2	73.6	190512	2	AC135699	AC135699 Rattus no
C 92	16.4	74.5	222746	2	EX510345	EX510345 Danio rer	C 165	16.2	73.6	190542	3	AC095015	AC095015 Drosophil
93	16.4	74.5	228802	2	AC023717	AC023717 Drosophil	C 166	16.2	73.6	199412	10	AC122342	AC122342 Mus muscu
94	16.4	74.5	251952	2	AC094693	AC094693 Rattus no	C 167	16.2	73.6	201415	2	AC117103	AC117103 Rattus no
C 95	16.4	74.5	281530	1	AP002550	AP002550 Escherich	C 168	16.2	73.6	202050	1	AL646065	AL646065 Ralstonia
C 96	16.4	74.5	290029	1	AE016978	AE016978 Shigella	C 169	16.2	73.6	203223	2	AC116077	AC116077 Rattus no
C 97	16.4	74.5	300409	1	AE016755	AE016755 Escherich	C 170	16.2	73.6	216062	2	AC130262	AC130262 Rattus no
C 98	16.4	74.5	308317	3	AE003487	AE003487 Drosophil	C 171	16.2	73.6	221118	2	AC095855	AC095855 Rattus no
99	16.4	74.5	323069	2	AC079737	AC079737 Homo sapi	C 172	16.2	73.6	223438	10	AL662804	AL662804 Mouse DNA
100	16.4	74.5	349980	6	AX044034	AX044034 Sequence	C 173	16.2	73.6	238055	2	AC131139	AC131139 Rattus no
101	16.2	73.6	462	10	RATATM2	M34138 Rat alpha-t	C 174	16.2	73.6	238055	2	AC131139	AC131139 Rattus no
C 102	16.2	73.6	753	6	BD246045	BD246045 Developme	C 175	16.2	73.6	239021	2	AC102973	AC102973 Rattus no
C 103	16.2	73.6	794	1	AX134487	AX134487 Providenc	C 176	16.2	73.6	240652	2	AC111587	AC111587 Rattus no
C 104	16.2	73.6	812	6	AX109605	AX109605 Sequence	C 177	16.2	73.6	240890	2	AC123328	AC123328 Rattus no
C 105	16.2	73.6	819	6	AX109606	AX109606 Sequence	C 178	16.2	73.6	245023	2	AC129467	AC129467 Rattus no
C 106	16.2	73.6	830	6	AX109604	AX109604 Sequence	C 179	16.2	73.6	245524	2	AC095456	AC095456 Rattus no
107	16.2	73.6	1050	10	RATMBRIA	M34135 Rat brain a	C 180	16.2	73.6	245654	2	AC116205	AC116205 Rattus no
108	16.2	73.6	1241	8	AY058156	AY058156 Arabidops	C 181	16.2	73.6	245907	2	AC095162	AC095162 Rattus no
C 109	16.2	73.6	1362	8	AC068616	AC068616 Oryza sat	C 182	16.2	73.6	251073	2	AC108534	AC108534 Rattus no
C 110	16.2	73.6	1366	3	AC006700	AC006700 Caenorhab	C 183	16.2	73.6	251244	2	AC103149	AC103149 Rattus no
111	16.2	73.6	1911	6	BD260923	BD260923 Novel met	C 184	16.2	73.6	252192	2	AC095221	AC095221 Rattus no
C 112	16.2	73.6	1911	6	AX036306	AX036306 Sequence	C 185	16.2	73.6	253021	2	AC103467	AC103467 Rattus no
113	16.2	73.6	1911	6	AX038916	AX038916 Sequence	C 186	16.2	73.6	253444	2	AC094541	AC094541 Rattus no
C 114	16.2	73.6	2488	10	RATMBRIA	M34136 Rat brain a	C 187	16.2	73.6	256703	2	AC114076	AC114076 Rattus no
C 115	16.2	73.6	3773	3	DNAC001656	DNAC001656 Drosophil	C 188	16.2	73.6	257071	1	AP002558	AP002558 Escherich
C 116	16.2	73.6	6022	6	AX251151	AX251151 Sequence	C 189	16.2	73.6	261089	2	AC094807	AC094807 Rattus no
C 117	16.2	73.6	6022	6	AX277916	AX277916 Sequence	C 190	16.2	73.6	261143	2	AC097135	AC097135 Rattus no
C 118	16.2	73.6	6022	6	AX323597	AX323597 Sequence	C 191	16.2	73.6	261434	2	AC130556	AC130556 Rattus no
C 119	16.2	73.6	6022	6	AX345974	AX345974 Sequence	C 192	16.2	73.6	270712	2	AC137360	AC137360 Rattus no
C 120	16.2	73.6	9888	6	AX346143	AX346143 Sequence	C 193	16.2	73.6	271725	2	AC116298	AC116298 Rattus no
C 121	16.2	73.6	10358	1	AE005403	AE005403 Escherich	C 194	16.2	73.6	290628	1	AE016983	AE016983 Shigella
C 122	16.2	73.6	10378	1	AE000275	AE000275 Escherich	C 195	16.2	73.6	290864	2	AC006702	AC006702 Caenorhab
123	16.2	73.6	10764	1	AE015166	AE015166 Shigella	C 196	16.2	73.6	300475	1	AE016761	AE016761 Escherich
C 124	16.2	73.6	10814	1	AE014472	AE014472 Brucella	C 197	16.2	73.6	303446	1	AE016941	AE016941 Bacteroid
C 125	16.2	73.6	11987	1	AE009465	AE009465 Brucella	C 198	16.2	73.6	304250	1	EX294152	EX294152 Pirellula
C 126	16.2	73.6	16668	6	BD246036	BD246036 Developme	C 199	16.2	73.6	308146	3	AE003673	AE003673 Drosophil
C 127	16.2	73.6	16784	7	AF513032	AF513032 Staphyloc	C 200	16.2	73.6	313735	2	AC094115	AC094115 Rattus no
C 128	16.2	73.6	16897	1	D90825	D90825 E.coli geno	C 201	16.2	72.7	998	11	CNS06F74	AL396038 T7 end of
C 129	16.2	73.6	18227	7	AF513033	AF513033 Staphyloc	C 202	16.2	72.7	9874	1	AE011350	AE011350 Leptospir
C 130	16.2	73.6	32179	3	AF047663	AF047663 Caenorhab	203	16.2	72.7	124215	8	AP002093	AP002093 Macaca mu
131	16.2	73.6	36579	2	AC134070	AC134070 Rattus no	204	16.2	72.7	151806	8	AC002093	AC002093 Oryza sat
C 132	16.2	73.6	56043	2	AC012761	AC012761 Drosophil	205	16.2	72.7	300050	1	AP004599	AP004599 Oceanobac
133	16.2	73.6	58464	2	AC110049	AC110049 Homo sapi	206	15.8	71.8	20	6	AX292024	AX292024 Sequence
C 134	16.2	73.6	74045	9	EX284671	EX284671 Human DNA	207	15.8	71.8	24	6	AX292391	AX292391 Sequence
C 135	16.2	73.6	80866	9	AC012649	AC012649 Drosophil	C 208	15.8	71.8	775	9	HS334076	HS334076 Homo sapi
C 136	16.2	73.6	81662	8	AB008265	AB008265 Arabidops	209	15.8	71.8	954	6	AR096464	AR096464 Sequence
137	16.2	73.6	87102	2	AL606983	AL606983 Danio rer	C 210	15.8	71.8	1101	6	AX459691	AX459691 Sequence
C 138	16.2	73.6	95968	3	CEY47D3B	AL031635 Caenorhab	211	15.8	71.8	1280	1	RHMNI-FHW	K01620 Rhizobium j

212 15.8 71.8 1281 6 E00713 Genomic DNA
213 15.8 71.8 1347 6 AR096466
214 15.8 71.8 1587 5 GG059421
215 15.8 71.8 1587 5 AR096465
216 15.8 71.8 2233 6 AR096463
217 15.8 71.8 2436 1 STYCTICA
218 15.8 71.8 3188 3 BT001752
219 15.8 71.8 3795 3 BT003565
220 15.8 71.8 5027 6 AR096462
221 15.8 71.8 8391 1 AY392415
222 15.8 71.8 11916 3 DME249115
223 15.8 71.8 13124 1 AE009105
224 15.8 71.8 13147 1 AE008070
225 15.8 71.8 13639 1 BS23937
226 15.8 71.8 19165 1 AC134163
227 15.8 71.8 22507 6 AX702457
228 15.8 71.8 27923 2 AC095226
229 15.8 71.8 35737 9 AC005263
230 15.8 71.8 39143 9 AC004490
231 15.8 71.8 49656 2 AC104957
232 15.8 71.8 53038 2 AC111637
233 15.8 71.8 60894 2 AC090930
234 15.8 71.8 62749 2 AC019842
235 15.8 71.8 74759 2 AC100661
236 15.8 71.8 77283 2 AC020099
237 15.8 71.8 78290 2 AC108631
238 15.8 71.8 83412 2 AC133611
239 15.8 71.8 89246 2 AC068226
240 15.8 71.8 90502 9 AC108676
241 15.8 71.8 93829 3 AC007122
242 15.8 71.8 95419 2 AC133704
243 15.8 71.8 98955 9 AC008886
244 15.8 71.8 99204 2 AC135098
245 15.8 71.8 101033 2 AC136177
246 15.8 71.8 110000 2 AC092921
247 15.8 71.8 110000 2 AC095226
248 15.8 71.8 110000 2 AC097907
249 15.8 71.8 110000 2 AC107271
250 15.8 71.8 110000 2 AC107271
251 15.8 71.8 110000 2 AC113863
252 15.8 71.8 110000 2 AC132690
253 15.8 71.8 110000 2 AC118411
254 15.8 71.8 110000 2 AC125991
255 15.8 71.8 110000 2 AC125991
256 15.8 71.8 110000 2 AC127799
257 15.8 71.8 110000 2 AC129112
258 15.8 71.8 110000 2 AC132558
259 15.8 71.8 110000 2 LMFLCHR16_03
260 15.8 71.8 114134 9 AF236876
261 15.8 71.8 115525 9 AC117491
262 15.8 71.8 115597 2 AC136166
263 15.8 71.8 118994 9 AL139823
264 15.8 71.8 121414 2 AC017234
265 15.8 71.8 128561 2 AC068729
266 15.8 71.8 129957 9 HS415G2
267 15.8 71.8 137061 2 AC136046
268 15.8 71.8 138719 2 AC126534
269 15.8 71.8 143897 9 AC011647
270 15.8 71.8 144552 8 CNS08CBX
271 15.8 71.8 151211 2 AC117337
272 15.8 71.8 151321 2 AC136264
273 15.8 71.8 152664 2 AC128855
274 15.8 71.8 155592 2 AC140736
275 15.8 71.8 157419 2 AC025389
276 15.8 71.8 160287 2 AC124150
277 15.8 71.8 161177 2 AC125850
278 15.8 71.8 163308 2 AC137174
279 15.8 71.8 165051 9 AC026950
280 15.8 71.8 165190 3 AC093096
281 15.8 71.8 165304 2 AC135746
282 15.8 71.8 165765 2 AC117979
283 15.8 71.8 166504 9 AC008275
284 15.8 71.8 167243 2 AC125956

285 15.8 71.8 167841 2 AC125767
286 15.8 71.8 169226 2 AC073309
287 15.8 71.8 169835 9 AC020896
288 15.8 71.8 170040 2 AC141975
289 15.8 71.8 173165 3 AC099005
290 15.8 71.8 173783 2 AC117135
291 15.8 71.8 173965 9 AL159977
292 15.8 71.8 175091 2 AC141991
293 15.8 71.8 176024 2 CNS09SC3
294 15.8 71.8 177109 2 AC118161
295 15.8 71.8 177278 3 AC007726
296 15.8 71.8 177787 9 AC133961
297 15.8 71.8 178778 2 AC105810
298 15.8 71.8 179115 2 AC097862
299 15.8 71.8 180000 1 AF222012
300 15.8 71.8 180290 2 AC123434

RESULT 1
LOCUS MBO298910/c 335 bp DNA linear BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain 1503/96.
ACCESSION AJ298910
VERSION AJ298910.1 GI:16416365
KEYWORDS repetitive element; tandem repeat; VNTR.
SOURCE Mycobacterium bovis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1
AUTHORS Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE Fluorescent VNTR (fVNTR) for improved high throughput molecular typing of Mycobacterium bovis
JOURNAL Unpublished
2 (bases 1 to 335)
AUTHORS Llewellyn-Hughes, S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial Diseases, Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT15 3NB, UNITED KINGDOM

FEATURES
source
1..335
/organism="Mycobacterium bovis"
/mol_type="genomic DNA"
/strain="1503/96"
/db_xref="taxon:1765"
repeat_region
73..127
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"
repeat_region
128..206
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"
repeat_region
207..292
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"

ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 8.1; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY 1 CCGGGCAAAACCTCGCCTAAC 22
|||||
Db 317 CCGGGCAAAACCTCGCCTAAC 296

RESULT 2
MBO298911/c 335 bp DNA linear BCT 23-OCT-2001
LOCUS

```

DEFINITION   Mycobacterium bovis VNTR, strain 4705.
ACCESSION    AJ298911
VERSION      AJ298911.1 GI:16416366
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium bovis
ORGANISM     Mycobacterium bovis
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 335)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..335
              /organism="Mycobacterium bovis"
              /mol_type="genomic DNA"
              /strain="4705"
              /db_xref="taxon:1765"
              73..127
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              128..206
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              207..292
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              repeat_region
              repeat_region
              repeat_region
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGCCTAAC 22
|||||
Db 317 CCGGCAAAACCTCGCCTAAC 296
|||||

RESULT 3
MMI298907/c
LOCUS          MMI298907
DEFINITION    Mycobacterium microti VNTR, strain 4777/97.
ACCESSION     AJ298907
VERSION       AJ298907.1 GI:16416362
KEYWORDS      repetitive element; tandem repeat; VNTR.
SOURCE        Mycobacterium microti
ORGANISM      Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 390)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              repeat_region
              repeat_region
              repeat_region
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGCCTAAC 22
|||||
Db 372 CCGGCAAAACCTCGCCTAAC 351
|||||

RESULT 4
MMI298909/c
LOCUS          MMI298909
DEFINITION    Mycobacterium microti VNTR, strain 4700/97.
ACCESSION     AJ298909
VERSION       AJ298909.1 GI:16416364
KEYWORDS      repetitive element; tandem repeat; VNTR.
SOURCE        Mycobacterium microti
ORGANISM      Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 390)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     source
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              repeat_region
              repeat_region
              repeat_region
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGCCTAAC 22
|||||
Db 372 CCGGCAAAACCTCGCCTAAC 351
|||||

```

LOCUS MBO298905 414 bp DNA linear BCT 23-OCT-2000

DEFINITION Mycobacterium bovis VNTR, strain AN 5.

ACCESSION AJ298905

VERSION AJ298905.1 GI:16416360

KEYWORDS repetitive element; tandem repeat; VNTR.

SOURCE Mycobacterium bovis

ORGANISM Mycobacterium bovis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G. Fluorescent VNTR (fVNTR) for improved high throughput molecular typing of Mycobacterium bovis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 414)

AUTHORS Llewellyn-Hughes, S.

TITLE Direct Submission

JOURNAL Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial Diseases, Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT15 3NB, UNITED KINGDOM

FEATURES

source 1..414

repeat_region 73..127

repeat_region 128..206

repeat_region 207..292

ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 414;

Best Local Similarity 100.0%; Pred. No. 8.1;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22

DB 396 CCGGCAAAACCTCGGCTAAC 375

MTU298906 469 bp DNA linear BCT 23-OCT-2000

LOCUS MTU298906

DEFINITION Mycobacterium tuberculosis VNTR, strain H37Rv.

ACCESSION AJ298906

VERSION AJ298906.1 GI:16416361

KEYWORDS repetitive element; tandem repeat; VNTR.

SOURCE Mycobacterium tuberculosis

ORGANISM Mycobacterium tuberculosis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G. Fluorescent VNTR (fVNTR) for improved high throughput molecular typing of Mycobacterium bovis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 469)

AUTHORS Llewellyn-Hughes, S.

TITLE Direct Submission

JOURNAL Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial Diseases, Veterinary Laboratories Agency, New Haw, Addlestone, Surrey, KT15 3NB, UNITED KINGDOM

FEATURES

source 1..469

LOCUS Loranxin-Mycobacterium tuberculosis

AE007121/c
 LOCUS 10894 bp DNA linear BCT 27-APR-2001
 DEFINITION Mycobacterium tuberculosis CDC1551, section 207 of 280 of the complete genome.
 ACCESSION AE007121 AB000516
 VERSION AE007121.1 GI:13882769
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis CDC1551
 Mycobacterium tuberculosis CDC1551
 Bacteria; Actinobacterii; Actinobacteridae; Actinomycetales;
 Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 10894)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
 Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
 Unpublished
 2 (bases 1 to 10894)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
 Direct Submission
 Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
 FEATURES
 source
 1..10894
 /organism="Mycobacterium tuberculosis CDC1551"
 /mol_type="genomic DNA"
 /strain="CDC1551"
 /db_xref="taxon:8331"
 /note="clinical strain"
 complement(247..1515)
 /gene="WT2989"
 complement(247..1515)
 /gene="WT2989"
 /note="similar to PID:1237016 SP:P51835 GB:AL009126; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="signal recognition particle FtsY"
 /protein_id="AAK47316.1"
 /db_xref="GI:13882770"
 /translation="MWELMTATAVIAALVIAALTLGLVLRRLRSLSPRPGVW
 DRSGYTAASSGITSPTPTOPAEIRIDTSGLPVGGDDATVPRDAPKRTIADVHLPEPE
 PEPQAPVEPADAIAPGRLERLRGLASQNALGRLGLGGDLDEDSQDYED
 TLLVADGPATASVQLSRSLASGNVFEADARAVLRVLINELOPGMDRSIRALP
 HAGPSVLLVGVNGTKTTVTKGLRVIVDGRVVGAAFTFRAAAADQLOQTWAAR
 VGAAVVRPEGADPASVAFADVDKGIAGADVWIDTAGRLHTKVGIMDELDKRVV
 TRRASVDELVLVDITQGLAQARVFAEVVDISGAVLTKLDGTAKGGIVFRVQDEL
 GVPVKVLGSGPDDLPFPFAFVDALLG"
 complement(1565..5182)
 /gene="WT2990"
 complement(1565..5182)
 /gene="WT2990"
 /note="similar to PID:1237015 SP:P51834 GB:AL009126; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="chromosome segregation SMC protein, putative"
 /protein_id="AAK47317.1"
 /db_xref="GI:13882771"
 /translation="MYLKSUTLKGFKSFAPTTLRPPGTVAVVNGSGKSNVVDAL
 AWMGEGQATLKGKMDVIFATSSRAPLGRAEVTVSDNSDALPIETVEVSTIR
 WPFDRGASVEYINGSCRLMDVQLSDSGPLGEMHVIVQGKLEILASRPEDRAF
 IEAAGVLKRRKRLKLDLTMANLRLTDLTTLRQLPLGRQAAQRAAI
 QADLRDLRLAADLVSRAREAVFQAEANMERHDEAAELAVASELAHSAV
 AEJSTRASIQHTWFGLSALAEVDAVTRIASRAHLDIEPVANSDTPRKPELEA

EAQVAVAEQQLLAELDARAEILDARAEELADREAEAEADRAHAAVTEAEADREGL
 ARLAGVETHARVESIDESVARLSERIEDDAAMRAQOTRAEFETVQGRIGELDQGEVH
 LDEHHTVAARLADREVAELQSAERQVARSURRIDALAVGLQSRDGAAMLAH
 NRSAGLFGSIAQLVKVRSYEAALAAALGPADALAVDGLTAAGSAVSAKQADGGR
 AVLSDWPAPQASGEMLPSCQAQWALDVLVEPPQLVGAMISLGVAVVNDLTE
 AMGLVIRPELRVAVTDGLVGAVSGGSDRLSTLEVTSIDKARSELAAAEALAA
 QLNAAAGALTEQARQDAEAQALANESDTAISMYBOLGRIGQEARAEENRL
 LOOTFEAVRTQTLDDVTQLETKAQETORVOAOPIDRQALISAADARAEVVE
 ARLAVTAERANAVRGADSLRRRAAEAREARVAAQARARLARHAAVAVAACGR
 LLAGRUHRAVDGASQLRDASACQORLAAMAAVDEVNTLSARLGELTDSUHRDELA
 NQAALURIQLEQVLEQFCMAADLITTEYHPHALPPTTELEMAFEQAREGEVIA
 PAFMPDRTVQERRAKREARALAEGRVPALEEFALAEERFNLSTQLEDEKARK
 DLGVADVADARILQVFNDAFVDEVEREFGVFTALPGGEGRLRLTPEDDMLTTGIEV
 EARPPCKTITRLSLGSGKALTAVAMLVAFRAPSFPYINDEVAALDDVNLRRLL
 SIPEQLREQSQIIITHQKPTMEVADALVGTVMQNDGITAVISQRMRGQOVDQVLTNS
 S
 complement(5194..5475)
 /gene="WT2991"
 complement(5194..5475)
 /gene="WT2991"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="acylphosphatase"
 /protein_id="AAK47318.1"
 /db_xref="GI:13882772"
 /translation="MSAPDVRLTAWHGVQGVGFWRWTRCRALGLGLTGYAANHADG
 RVLVVAQGPRAACQKLLOLQDQTTTPGRVAKVVDWSQSTEQITGFSER"
 complement(5462..5875)
 /gene="WT2992"
 complement(5462..5875)
 /gene="WT2992"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK47319.1"
 /db_xref="GI:13882773"
 /translation="MTQLWERTGTTRYIGRSTRGAQVLVGSDEVGVFTPEGLKTA
 LAACSMASDOPLARLDGDDYQAVVKSADRDQERYPLIETMELDSLGLTDEKE
 RLVLVIRAVELACTVGRTLKSGTTNLEVVVDGA"
 5867..6013
 /gene="WT2993"
 5867..6013
 /gene="WT2993"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK47320.1"
 /db_xref="GI:13882774"
 /translation="WRRGFFLPGLCVEDIGSAPARESIALTSIMROHPYFRGRGW
 HRGQ"
 complement(5977..6846)
 /gene="WT2994"
 complement(5977..6846)
 /gene="WT2994"
 /note="similar to GP:3150225; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="formamidopyrimidine-DNA glycosylase"
 /protein_id="AAK47321.1"
 /db_xref="GI:13882775"
 /translation="WPELPEVVRGILQAVTGTITETVTHPRVHRHDGAPDL
 TARLKGARINGDRGKTLMLTILTAGVHRRTDIALVHLGMSQMLLGAVPCAHRV
 ISALLDDTGFADQRTFGGILLADLVTVDSGVVPVFAHLARDPLDPRFDCDAVR
 VLRRKSLKQLLDQVYSGIGNIYADEALVRKAVKVAHVAATLRCRRLGAVLHAAA
 DVMREAAKAGGTSFDSLTVYVNGSGYFERSLDYAGEGNCRCRCGAVIRERFMRS
 SFYCPRCQPRRK"
 complement(7205..7927)
 /gene="WT2995"
 complement(7205..7927)

SOURCE
ORGANISM
Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
AUTHORS
1. Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and
Hewinson, G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 (Microbiology)
2 (bases 1 to 318050)

REFERENCE
AUTHORS
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail: tgarnier@pasteur.fr Submitted on behalf of The
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. .318050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
/complement(99..875)
/locus_tag="Mb2882c"
/complement(99..875)
/locus_tag="Mb2882c"
/EC_number="1.-.-.-"
/note="Mb2882c, - len: 258 aa. Equivalent to RV2857c;
len: 258 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 258 aa overlap). Probable short-chain
dehydrogenase/reductase [EC 1.-.-.-], highly similar to
various dehydrogenases e.g. O80668|SC135.33c PROBABLE
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
identity in 253 aa overlap); O91376|PA1649 from
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
(253 aa), FASTA scores: opt: 569, E(): 2.e-28, (39.2%
identity in 255 aa overlap); Q9EX74|MLHA SDR-LIKE ENZYME
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);
etc. Also similar to many Mycobacterium tuberculosis
dehydrogenases e.g. FBG3|RV2002|MT2058|MCY39.16c
PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
DEHYDROGENASES/REDUCTASES (SDR) FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE SHORT-CHAIN TYPE
DEHYDROGENASE/REDUCTASE"
/protein_id="CAD96369.1"
/db_xref="GI:31619629"
/translation="MMDLSQLRAGRVAVITGGSGIGLAAGRRNRAGATIVVGDVDF
EAGCAADSLGIFVFTVDCEDAVNGLFDGAEATITAFNNGIASPPEDNLLEN
TELAQWQVDNLKSVYLCRAALRHVVLGAGTSVNTASFAVNVGSGATSYTAS
KGVGLASRELGVQFARQGRVNLCPGPNVTELLQELFAKNPERAARMVHVPGLGRF
AEPEETAARVAFSLDASGLTATFLVDGGISSAVTPL"
/complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
/complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"

FEATURES
source

gene

CDS

gene

CDS

```

/EC number="1.2.1.3"
/Note="Mb2883c, aldC, len: 455 aa. Equivalent to Rv2858c,
len: 455 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 455 aa overlap). Probable aldC,
aldehyde dehydrogenase (EC 1.2.1.3). Similar to many e.g.
O88069|SCI35.34c PUTATIVE ALDEHYDE DEHYDROGENASE from
Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,
E(): 6.4e-109, (64.5% identity in 448 aa overlap);
O9FAB1|ALDH1|BT-ALDH ALDEHYDE DEHYDROGENASE from Bacillus
thermoleovorans (497 aa), FASTA scores: opt: 1157, E():
2.1e-64, (44.3% identity in 458 aa overlap); O33455|CYMC
P-CUMIC ALDEHYDE DEHYDROGENASE from Pseudomonas putida
(494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15%
identity in 452 aa overlap);
P40047|DHA5 YEAST|ALDS|ALDH5|ALD3|YERO73W ALDEHYDE
DEHYDROGENASE from Saccharomyces cerevisiae (Baker's
Yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60,
(38.55% identity in 459 aa overlap);
P80668|PEAB-ECOLI|PADA|MAOB|B1385 PHENYLACETALDEHYDE
DEHYDROGENASE (EC 1.2.1.39) from Escherichia coli strain
K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2%
identity in 462 aa overlap); etc. Also similar to many M.
tuberculosis dehydrogenases e.g. P71823|Rv0768|MTCY369.13
(489 aa), FASTA score: (38.1% identity in 467 aa overlap).
Contains P800687 Aldehyde dehydrogenases glutamic acid
active site and P800070 Aldehyde dehydrogenases cysteine
active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES
FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"
/protein_id="GI:31619630"
/db_xref="GI:31619630"
complement(2236..3162)
/locus_tag="Mb2884c"
complement(2236..3162)
/locus_tag="Mb2884c"
/EC number="2.-.-.-"
/EC number="6.3.5.-"
/Note="Mb2884c, -, len: 308 aa. Equivalent to Rv2859c,
len: 308 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 308 aa overlap). Possible
amidotransferase (EC 6.3.5.- or 2.-.-.-), equivalent (but
longer 58 aa) to Q9CBU9|ML1573 POSSIBLE AMIDOTRANSFERASE
from Mycobacterium leprae (249 aa), FASTA scores: opt:
1226, E(): 3e-64, (71.53% identity in 239 aa overlap).
Also similar to other amidotransferases and hypothetical
proteins, but shorter in N-terminus e.g. O88072|SCI35.37
HYPOTHETICAL 25.3 KDA PROTEIN from Streptomyces coelicolor
(242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65%
identity in 235 aa overlap); AAQ79730|Q97188|Cac1764
PREDICTED GLUTAMINE AMIDOTRANSFERASE from Clostridium
acetobutylicum (241 aa), FASTA scores: opt: 458, E():
1.6e-19, (32.95% identity in 246 aa overlap);
AAK75201|Q97QV9|SP1089 GLUTAMINE AMIDOTRANSFERASE CLASS I
from Streptococcus pneumoniae (229 aa), FASTA scores: opt:
431, E(): 5.6e-18, (34.75% identity in 236 aa overlap);
etc. Contains three 17 aa repeats at the N-terminus very
similar to those in other Mycobacterium tuberculosis
proteins e.g. Q10699|YJ30 MYCTU|Rv2090|MT2151|MTCY49.30
PUTATIVE 5'-3' EXONUCLEASE Rv2090 (EC 3.1.11.-).
/codon_start=1
/transl_table=11
/product="POSSIBLE AMIDOTRANSFERASE"

```

```

/protein_id="CAD96571.1"
/db_xref="GI:31619631"
/translations="MOLASRSDDGDPRLRSPVSDGDDPRLRSPASRLRSPV
DGDGDPRLRSPASRLRSPVGLTAYLEQVTRGVWDIPAGYTPADYFEIGITVAGG
VAVLLPQVPDPESVGVCLDSHALVITGQYLDPAAYGQEPHPTDPRGRDAWEF
ALLRGLAQMGVLCRGTVHNLVALGQDILGHLPDLILGHSHGAGNVFRLPVHT
ASTRIAEILIGESADVCPTVHOAIDOVGGLVAVSDVDVIEALELPGDTFTVLAQW
HPEKSDDLLEKALVDRAAGVAGRSQAEPR"
complement(3143..4516)
/genes="glnA4"
/locus_tag="Mb2885c"
/EC number="6.3.1.2"
/Note="Mb2885c, glnA4, len: 457 aa. Equivalent to Rv2860c,
len: 457 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 457 aa overlap). Probable glnA4,
glutamine synthetase class II (EC 6.3.1.2), similar to
many glutamine synthetases e.g. O88070|SCI35.35c from
Streptomyces coelicolor (462 aa), FASTA scores: opt: 1947,
E(): 8.2e-120, (64.15% identity in 452 aa overlap);
Q98H15|MLL3074 from Rhizobium loti (Mesorhizobium loti)
(465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7%
identity in 452 aa overlap); Q98EM0|MLL4187 from Rhizobium
loti (Mesorhizobium loti) (456 aa), FASTA scores: opt:
698, E(): 4.6e-38, (33.5% identity in 454 aa overlap);
Q9CDL9|GLNA from Lactococcus lactis (subsp. lactis)
(Streptococcus lactis) (446 aa), FASTA scores: opt: 633,
E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.
Also similar to three other potential glutamine synthetases
in Mycobacterium tuberculosis:
Q10378|GLN2 MYCTU|GLNA2|Rv2222c|MT2280|MTCY190.33c|MTCY427
.03c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score:
(31.1% identity in 453 aa overlap); Rv1878|glnA3 and
Rv2220|glnA1. BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE GLUTAMINE SYNTHETASE GLNA4 (GLUTAMINE
SYNTHASE) (GS-II)"
/protein_id="CAD96572.1"
/db_xref="GI:31619632"
/translations="MTGRSPPLAWTELERLVAAGDVDTIVAFVTDMOGLRAGKRSIG
RHFVDDIATRGVECCSYLLAVDVLNTVPGYAMASWDIYGDMVTNPDLSTRLPWL
PGTALVIALVWADGSEVAVSPRSILRRQDLKARGLVADVATELEIFVDPYRQA
WASGYGLTPADSYNDIDYAILASSMEPELRDLRLGNAAGLFEAVKGCNMGQGEI
GFYDEALVTCNDHAIYKNGAKEIADQHKSLITFMKYDEREGNSCHTIVSLRGTDGS
AVFADSNPGHGMSSMFRSVAGQLATREFTLCYAPTINSYKRFADSSFAFTALAWGL"
Query Match 100.0%; Score 22; DB 1; Length 318050;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCCGCCTAAC 22
|||||
DB 71422 CCGGGCAAAACCTCCGCCTAAC 71401
|||||
RESULT 10
BX842581/c
ID BX842581
XX
AC BX842581; AL008883; AL008967; AL021070; AL021287; AL021309; AL123456;
AC Z74024; Z74697; Z81331; Z83018; Z83857; Z83866; Z83866; Z83866; Z83866;
SV BX842581.1
XX
DT 21-NOV-2003 (Rel. 77, Created)
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)
XX
DE Mycobacterium tuberculosis H37Rv complete genome; segment 10/13
XX complete genome.
XX

```

XX	Myxobolium tuberculosis H37Rv		
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;		
OC	Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.		
XX			
XX	[1]		
RN	MEDLINE; 98295987.		
RX	PUBMED; 9634230.		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eigmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,		
RA	Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,		
RA	Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,		
RA	Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,		
RA	Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,		
RA	Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RA	Deciphering the biology of Mycobacterium tuberculosis from the complete		
RT	genome sequence;		
RT	Nature 393:537-544 (1998).		
XX			
XX	[2]		
RN	PUBMED; 12368430.		
RX	Camus J.C., Pryor M.J., Medigue C., Cole S.T.;		
RA	"Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";		
RL	Microbiology 148:2967-2973 (2002).		
XX			
XX	[3]		
RN	1-348676		
RP	Parkhill J.;		
RA	Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.		
RL	Submitted on behalf of the Mycobacterium tuberculosis sequencing and		
RL	mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,		
RL	Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut		
RL	Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:		
RL	parkhill@sanger.ac.uk		
XX			
CC	Notes:		
CC	Details of M. tuberculosis sequencing at the Sanger Centre		
CC	are available on the World Wide Web.		
CC	(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)		
XX			
XX	Key	Location/Qualifiers	
PH	source	1..348676	
FT	/db_xref="taxon:8332"		
FT	/mol_type="genomic DNA"		
FT	/organism="Mycobacterium tuberculosis H37Rv"		
FT	/strain="H37Rv"		
FT	complement(148..183)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(148..4539)		
FT	/note="4392 bp direct repeat region"		
FT	complement(222..257)		
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(298..333)		
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(374..409)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(447..482)		
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(519..554)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(590..625)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1664..1699)		

FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(740..775)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(811..846)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(884..919)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(958..993)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1031..1066)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1104..1139)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1176..1211)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1248..1283)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1322..1357)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1396..1431)		
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to		
FT	sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1467..1486)		
FT	/note="20 bp partial direct repeat, CCCCAGAGGGGACGGAAC,		
FT	of sequence GTCGTACAGACCCCAAAACCCCGAGAGGGGACGGAAC"		
FT	complement(1486..2860)		
FT	/note="IS6110-11, len: 1375 bp. Insertion sequence IS6110."		
FT	/insertion seq="IS6110-11"		
FT	complement(1529..2467)		
FT	/note="RV2814c, (WTCY16B7.29), len: 312 aa. Probable		
FT	transposase, highly similar to others e.g.		
FT	P37137 RV0796 MTV042.06 PUTATIVE TRANSPOSASE FOR INSERTION		
FT	SEQUENCE ELEMENT IS986/IS6110 from Mycobacterium		
FT	tuberculosis (328 aa), FASTA scores: opt: 2103, E():		
FT	6.1e-132, (100.0% identity in 312 aa overlap); etc. Start		
FT	unlikely."		
FT	/transl_table=11		
FT	/function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION		
FT	ELEMENT IS6110."		
FT	/locus tag="RV2814c"		
FT	/product="PROBABLE TRANSPOSASE"		
FT	/protein_id="CA03675.1"		
FT	/translation="LITREIADHOGHREGPDGLRWVESICTQLTELGVPIAPSTVYDH		
FT	INREPSRELRDGLKEHLSRSHVHAANYGYGARKVNLINREGISVARTVERLMTKL		
FT	LSGTRGKARRITADPAPADLVQRRFGPPAPNRLWADLTVSTWAGYAVFT		
FT	DAYARILGWRVASTMATSMLDAIEQAIWTRQEQGLVDKDVHHTDRGQVTSIRFS		
FT	ERLAAAGIOPSGVAGSSYDNLAEITNGLYKTELKPGKWRSDVELATARWVDWF		
FT	NHRLVQVCGDVPVPELEAAVYVQRPAAG"		
FT	complement(2464..2790)		
FT	/note="RV2815c, (WTCY16B7.28), len: 108 aa. Probable		
FT	transposase, identical from aa 51 with P19772 Y12_MYCTU		
FT	PUTATIVE TRANSPOSASE (INSERTION ELEMENT IS986) from		
FT	Mycobacterium tuberculosis (59 aa), FASTA scores: opt: 365,		
FT	E(): 1.1e-19, (96.6% identity in 59 aa overlap); and other		
FT	transposases."		
FT	/transl_table=11		
FT	/function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION		
FT	ELEMENT IS6110."		
FT	/locus tag="RV2815c"		
FT	/product="PROBABLE TRANSPOSASE"		
FT	/protein_id="CA03633.1"		
FT	/translation="MSGGSSRRYPPELRERAVRMVAIRGQHDSEWAAISEVARILGVG		
FT	CAETVKRWVQAVDAGAPGTTTESBELKRLRDNAELRRANILKTASAFFAELD		

```

FT repeat_unit complement(2845..2860)
FT /note="16 bp partial direct repeat, GTGCTGACGCCAAA, of
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(2901..2936)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(2975..3011)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3049..3084)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3121..3156)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3193..3228)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3266..3301)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3338..3373)
FT /note="36 bp direct repeat, 32 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3399..3434)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3475..3511)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3548..3583)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3624..3659)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3701..3736)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGACGGAAC"
FT complement(3774..3809)

Query Match 100.0%; Score 22; DB 15; Length 348676;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCCTAAC 22
Db 120812 CCGGCAAAACCTCGCCTAAC 120791

RESULT 11
AC095431/c LOCUS
DEFINITION Rattus norvegicus clone CH230-7J16, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC095431
VERSION AC095431.11 GI:30467895
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 242139)
Munzy D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, X., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

```

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Geisregeris, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensheva, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Paokelme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
 Sanders, W., Savery, G., Schefer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shwartzbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 242139)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 242139)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:24941182.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

Direct Submission
Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(<http://fruitfly.berkeley.edu>) or send email to
cdna@fruitfly.berkeley.edu.

```
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..2629
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
1..2629
/gene="Sphk1"
/db_xref="FLYBASE:FBgn0030300"
235..2160
/gene="Sphk1"
```

```

235..2160
/gene="Sphk1"
/note="Longest ORF"
/codon_start=1
/product="RB64552p"
/protein_id="AAW50306.1"
/db_xref="GI:21429174"
/db_xref="FLYBASE:FBgn0030300"
/translation="VTANTGSGENLMGNKGKTHEPTSTSTDEVASGTFTSEISFF

```

VNSRKNOSIKITQVKLCPEGVYTRRETEDEDDHINEQLIRIDDIIGSRYGRRLLKRRARV
 GLNSCRNPNQGEASDPDSNSALVIYAVLKKPKLRVRSOTLIRFRSSNDYGV
 NILOTAEMGHHTIRKHKRNGSSPADCKQGLLILNPKSGSGKGRLLFOKQVAPLITVE
 AELVQDILITTHPOVAKFVRPRRDLITRYSIGIVVASGDGLFYEVNLGLMRWDREA
 CREPLDGLITPGSGNGHKLAKSVAAHCNHPVEPPIILHATITCMACKSTPMNVVVELAT
 RKHFVPMYFVLSVGLGLADIDIESRLSISGAQRFTLWAKLIGURSTKGRVSYLLS
 GKKGKFPVYEAARELPAPSTADIGIRSLPNAGEFDLPBEEGEVADLQGEQFADALS
 LDRSVYRQHDSSWHSAMSRRTAYISGLVPSMSRNSRMGISOTIRAEANAFAFRVPTG
 TTPPLQMLSDSGWICEDGPFVWHAAYTHLSSDVPFAPESRLDDGLLYLVYIRRG
 VSRHOLLNMLNLNAGTHLPIGEDPFIKVPCKRAPRIEPSSSDGLLVWDGGERVEYGPFI
 QAEVYVPLINVTWITSQ"

ch 80.9%; Score 17.8; DB 3; Length 2629;
 l similarity 90.5%; Pred. No. 8.7e+02;
 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 CGGGGCAAAACCTCCGCCTAA 21
 |||||
 26 CGGGGCATAACTCCGCCTGA 2106

AC013897 46594 bp DNA linear HTG 16-NOV-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 AC013897
 AC013897.1 GI:6437438
 HTG: HTGS_PHASE2
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 46594)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT
This sequence was identified as CDM:10210473 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..46594
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 80.9%; Score 17.8; DB 2; Length 46594;
Best Local Similarity 90.5%; Pred. No. 7.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
0;

QY 1 CCGGGCAAAACCTCGCCCTAA 21
|||||
DB 42642 CCGGGCATACCTCGCCCTGA 42662
|||||

RESULT 14
AC023718/c
LOCUS
DEFINITION
Drosophila melanogaster X BAC RP98-7P15 (Roswell Park Cancer Institute Drosophila BAC Library) complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC023718
AC023718.4 GI:21397192
HTG.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS

Munzy, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A., Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H., Dugan-Rocha, S., Sodergren, E.S., Hodgson, A.H., Chen, R.C., Ayele, M., Scott, G.S., Worley, K.W., Amaratunga, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C., Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K., Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E., Draper, H., Emery-Cohen, A., Ferreira, S., Garg, N.D.S., Houck, J., Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C., Liu, W., Mattel, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B., Nelson, K.A., Ndaasa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V., Scheeler, P., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M., Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O., Wheeler, D., Weinstein, G., Gibbs, R. and Venter, J.C.

Direct Submission
Unpublished
2 (bases 1 to 164415)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Binage, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Duggan-Rocha, S., Durbin, K.J., Earnhart, C., Edgall, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegued, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.D., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Unpublished
3 (bases 1 to 164415)
Worley, K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

4 (bases 1 to 164415)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Binage, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Duggan-Rocha, S., Durbin, K.J., Earnhart, C., Edgall, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegued, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.D., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
 Submitted (18-DEC-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 164415)
 Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Banks, T., Barbacia, J.,
 Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louisedge, H., Lozardo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,
 Sisson, I., Sodergren, E., Sonaiké, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Weinstock, G. and Gibbs, R.

Query Match 80.9%; Score 17.8; DB 3; Length 164415;
 Best Local Similarity 90.5%; Pred. No. 7.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCCGCTAA 21
 Db 78760 CCGGCATAACTCCGCTGA 78740

RESULT 15
 AC102929/c
 LOCUS
 DEFINITION
 AC102929
 AC102929.2 GI:30017624
 HTG: HTGS PHASE1; HTGS DRAFT.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 166296)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-226P13
 Unpublished
 2 (bases 1 to 166296)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckalter, B.,
 Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,
 Chospel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
 Cooke, P., DeAcellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Menus, L., Mihova, T., Mienga, V., Murphy, T., Navly, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
 Oliver, C., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topf, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Vielh, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 166296)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melgrim, J., Meneva, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi:17061039.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20088
Center clone name: 226_P_13

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1
* 1312: contig of 1312 bp in length
* 1313 1412: gap of 100 bp
* 1413 10111: contig of 8699 bp in length
* 10112 10211: gap of 100 bp
* 10212 21544: contig of 11333 bp in length
* 21545 21644: gap of 100 bp
* 21645 74138: contig of 52494 bp in length
* 74139 74238: gap of 100 bp
* 74239 110959: contig of 36721 bp in length
* 110960 111059: gap of 100 bp
* 111060 166296: contig of 55237 bp in length.

```

FEATURES

```

Source
1..166296
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-226P13"
/clone_lib="RPCI-24 Male Mouse BAC"

1..1312
/misc_feature
/feature="assembly_fragment"
1413..10111
/misc_feature
/feature="assembly_fragment"
10112..21544
/misc_feature
/feature="assembly_fragment"
21645..74138
/misc_feature
/feature="assembly_fragment"
74239..110959
/misc_feature
/feature="assembly_fragment"
111060..166296
/misc_feature
/feature="assembly_fragment"

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 166296;
Best Local Similarity 90.5%; Pred No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 2 CGGCAAAACCTCCGCTTAC 22

Db 89309 CGAGCAAAACCGCGCTTAC 89289

RESULT 16

AC117358
LOCUS
DEFINITION
AC117358
VERSION
KEYWORDS
SOURCE
ORGANISM

AC117358 229648 bp DNA linear HTG 19-NOV-2002
Rattus norvegicus clone CH230-307G20, WORKING DRAFT SEQUENCE.
AC117358
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 229648)

AUTHORS

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Y., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, N., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pflanz, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 229648)

AUTHORS

Worley, K.C.

Direct Submission

TITLE

JOURNAL

Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 229648)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23811708.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUNO
 Center clone name: CH230-307G20
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 22100 bases at least Q40
 Consensus quality: 22716 bases at least Q30
 Consensus quality: 223928 bases at least Q20
 Estimated insert size: 227432; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 229648: contig of 229648 bp in length.
 Location/Qualifiers
 1..229648
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-307G20"
 1..2698
 /note="wgs end extension
 clone_end:Sp6"
 3433..4170
 /note="clone boundary
 clone_end:Sp6"
 site:
 end sequence:BZ254694"
 183068..228499
 /note="clone boundary
 clone_end:T7"
 site:
 end sequence:BZ254693"
 228594..229648
 /note="wgs end extension
 clone_end:T7"

ORIGIN
 Query Match 80.9%; Score 17.8; DB 2; Length 229648;
 Best Local Similarity 90.5%; Pred. No. 7.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGCGAAACCTCGCCTAA 21
 Db 210623 CCGGCGAAACATCCCTAA 210643
 RESULT 17
 AC097187
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-101L21, *** SEQUENCING IN PROGRESS
 ***, 3 unordered pieces.
 AC097187
 ACCESSION
 VERSION
 KEYWORDS
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 242397)
 Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, J.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusshewa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pul, L.,
 Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., S., Shen, H.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstein, G. and Gibbs, R. A.
 Direct Submission
 Unpublished

Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., Smith,H.O., Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 254540)
Worley,K.C.

Direct Submission
Submitted (26-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 254540)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21903160.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GWIK
Center clone name: CH230-221N23

Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 230554 bases at least Q40
Consensus quality: 235111 bases at least Q30
Consensus quality: 237406 bases at least Q20
Estimated insert size: 258735; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 244381: contig of 244381 bp in length
244382 244481: gap of unknown length
244482 244970: contig of 5389 bp in length
249871 249970: gap of unknown length
249971 251359: contig of 1389 bp in length
251360 251459: gap of unknown length
251460 253223: contig of 1764 bp in length
253224 253323: gap of unknown length
253324 254540: contig of 1217 bp in length.

Location/Qualifiers
1..254540
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-221N23"
1..1113
/note="wgs contig"
244482..254538

misc_feature
misc_feature

TITLE
JOURNAL
REFERENCE
AUTHORS
TAXID
JOURNAL
TITLE
JOURNAL
COMMENT

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 254540;
Best Local Similarity 90.5%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCTAA 21
|||||
Db 25401 CCGGGCAAAACATCCCCCTAA 25421
|||||

RESULT 19
AC115140
LOCUS
DEFINITION
Rattus norvegicus clone CH230-99N12, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
AC115140
AC115140.6 GI:23270043
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 260335)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Purn,C., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.B., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Deigado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,J., Harvey,Y., Haveland,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshaw,L., Loulseg,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwaeleleh,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Yoon,L., Yoon,V., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., Smith,H.O., Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 Reference 2 (bases 1 to 260335)
 Worley, K.C.
 Direct Submission
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 260335)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 23, 2002 this sequence version replaced gi:21736944.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNXU
 Center clone name: CH230-99N12
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 225253 bases at least Q40
 Consensus quality: 225495 bases at least Q30
 Consensus quality: 227152 bases at least Q20
 Estimated insert size: 239709; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 4 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 172711: contig of 172711 bp in length
 172712 172811: gap of unknown length
 172812 191948: contig of 19137 bp in length
 191949 192048: gap of unknown length
 192049 258860: contig of 66812 bp in length
 258861 258960: gap of unknown length
 258961 260335: contig of 1375 bp in length.
 Location/Qualifiers
 1..260335
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-99N12"
 1..1294
 /note="wgs contig"
 175748..175884
 /note="wgs_contig"

FEATURES
 source
 misc_feature
 misc_feature
 ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 260335;
 Best Local Similarity 90.5%; Pred. No. 7.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 CGGCGAACCCTCGGCTTAC 22
 Db 223788 CGGAGAAAACCTCTCTTAC 223808
 RESULT 20
 AC133723 286789 bp DNA linear HTG 20-NOV-2002
 LOCUS Rattus norvegicus clone CH230-107H7, *** SEQUENCING IN PROGRESS.
 DEFINITION *** 8 unordered pieces.
 AC133723
 VERSION AC133723.2 GI:25138891
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 286789)
 AUTHORS Muzny, D., Maric, Metzker, M., Lee, A., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowal, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusnewa, L., Louise, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwaekeme, O., Okwuonu, G., Olarpunsaogon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pflankoch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Weinstock, G. and Gibbs, R.A.
 Direct Submission

TITLE

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 286789)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (18-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 286789)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23096637.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KCIF
Center clone name: CH230-107H7
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 231208 bases at least Q40
Consensus quality: 237333 bases at least Q30
Consensus quality: 241696 bases at least Q20
Estimated insert size: 241840; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3691: contig of 3691 bp in length
* 3692 3791: gap of unknown length
* 3792 14444: contig of 10653 bp in length
* 14445 14544: gap of unknown length
* 14545 274434: contig of 259890 bp in length
* 274435 274534: gap of unknown length
* 274535 275544: contig of 1010 bp in length
* 275545 277634: contig of 1989 bp in length
* 277635 277734: gap of unknown length
* 277735 279878: contig of 2145 bp in length
* 279879 283039: gap of unknown length
* 283040 283140: contig of 3661 bp in length
* 283140 286789: contig of 3650 bp in length.
*
* Location/Qualifiers
* 1..286789
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-107H7"

FEATURES
source
misc_feature
1..1219
/note="wgs end extension
clone end:Sp6"
3792..5376
/note="wgs end extension
clone end:Sp6"
14545..15732
/note="wgs end extension
clone end:Sp6"
20525..21557
/note="wgs end extension
clone end:Sp6"
25203..182322
/note="clone boundary
clone end:Sp6
site:
end sequence:BH360481"
248024..249251
/note="wgs contig"
255512..258522
/note="wgs contig"

misc_feature
248024..249251
/note="wgs contig"
255512..258522
/note="wgs contig"

ORIGIN
Query Match 80.9%; Score 17.8; DB 2; Length 286789;
Best Local Similarity 90.5%; Pred. No. 7.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTTAA 21
Db 151363 CCGGGCAAAACATCCGCTTAA 151383

RESULT 21
AE003486/c
LOCUS
DEFINITION
Drosophila melanogaster chromosome X section 38 of 74 of the
complete sequence.
ACCESSION
AE003486 AE002593 AE014298
VERSION
AE003486.2 GI:228333082
KEYWORDS
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,X.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
Abril,J.P., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,B., Beasley,E.M.,
Beeson,K.Y., Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S.,
Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,
Burtis,K.C., Busam,D.A., Butler,H., Cadiou,E., Center,A.,
Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
Dietz,S.N., Dodson,K., Dunn,P., Dupuis,R.L., Ewing,M., Gabor,G.L.,
Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
Fierrea,S., Fleischmann,W., Fowler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Iqbal,Z., Iliash,I., Liang,Y., Lin,X., Liu,X., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Muzny,D.M., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,

```



```
/gene="dlgl"
/locus tag="CG1725"
/notes="dlgl gene product from transcript CG1725-RA"
/codon_start=1
/product="CG1725-PA"
/protein_id="AA48038.2"
/db_xref="GI:22833085"
/db_xref="FLYBASE:FBgn0001624"
/translation="MTTRKKRGDGGGGGFKKVSFLNLSLHKASSTKYNDDSW
LYEDIQERNGSLGFIAGCTDNPHGTDTSTYIKLISGGAADGRLSINDIIVS
INDVSVDVPHASNDALKAGNVKLVHVKRGTTATPAAGSAGDARDASAGPKV
IEDLVKGGKGLGFSIAGGIGNHIFGDNIGYITKMDGAAQVDELSTGDKLJAVR
TNGSEKLENTHELAVATLKSITDKVILIKTCHLTTSASGGGGGSLSSGQQLSQS
OSQATQSQSOHQHATPMVNSOSTGALNSGQTVDPSPISQAAAAAANAAS
ASASVIASNNITNTTITVTATATASNSSSKLPPSLGANSISISINSNSNSNNIN
```

Query Match 80.9%; Score 17.8; DB 3; Length 328128;
Best Local Similarity 90.5%; Pred. No. 7.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCCTAA 21
|||||
Db 75988 CCGGGCATAACCTCCGCCTGA 75968
|||||

RESULT 22
OSJN00127 163448 bp DNA linear PLN 14-NOV-2003
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0084K01,
DEFINITION complete sequence.
ACCESSION AL506999.3 GI:32489031
VERSION
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J.,
Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Tang, Y.,
Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L. S., Yu, Z., Fan, D.,
Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J.,
Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R.,
Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, G., Tu, Y., Jia, J.,
Zhang, Y., Chen, J., Wang, H., Chen, X., Shao, C., Sun, Y., Hu, Q.,
Zhang, X., Zhang, W., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S.,
Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, J.,
Li, J., Hong, G., Xue, Y. and Han, B.
Sequence and analysis of rice chromosome 4
Nature 420 (6913), 316-320 (2002)

2 22337377
12447439

3 Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q., Hu, X.,
Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J., Zhang, L.,
Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y., Li, C.,
Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K., Zhou, B.,
Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P., Fu, G.,
Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F., Tu, Y. F.,
Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Zhang, H., Chen, X. Y.,
Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W., Wang, L. J.,
Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L., Zhu, F. H. and
Hong, G. F.
Direct Submission
Submitted (26-SEP-2001) Han Bin, National Center for Gene Research,
Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn
Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJNBa0084K01.
On Jul 9, 2003 this sequence version replaced gi:21912497.
Web site: http://www.ncgr.ac.cn

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL

REMARK
COMMENT

----- Summary Statistics
Assembly program: phrap

This is a complete sequence.

Genes were identified by a combination of several methods: Gene
prediction programs including Fgenesh (<http://www.softberry.com/>),
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMark-ES (Sean
Bddy, <http://genemark.biology.gatech.edu/Genemark/>), tRNAscan-SE (Sean
Bddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the
complete sequence against NCBI none redundant protein database (nr)
(<http://ncbi.nlm.nih.gov/blast/db/>) and the EST database at NCGR.

Location/Qualifiers

FEATURES

source

1. 163448

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="4"
/clone="OSJNBa0084K01"
/clone_lib="CUGI-OSJNBa"
1269..1562
/gene="OSJNBa0084K01.1"
join(1269..1341,1435..1562)
/gene="OSJNBa0084K01.1"
/codon_start=1
/protein_id="CAE04829.1"
/db_xref="GI:32489032"
/translation="MYTRWSEKPGKILWLTGTAGIMITNVVTRVDMQKILQE
EDEAAAAAAMASGERVLKDD"
6124..9960
/gene="OSJNBa0084K01.2"
join(6124..6288,6370..6496,6568..6698,7052..7153,
7236..7319,7403..7630,7833..8020,8122..8226,8307..8630,
9445..9202,9298..9404,9607..9663,9751..9960)
/gene="OSJNBa0084K01.2"
/codon_start=1
/protein_id="CAE04830.1"
/db_xref="GI:32489033"
/translation="MLQVEALAAVLASAAAARAABAVPLHDALGLVLAEDVRAP
DLPPFRASIKGVAVASDGPXYVITESRAGDDALGVVTPGVAVVTCGPDP
GADVAVQVETQLAGAPGSKVRILVIRVTOGQDIRVNGCDIOKDSVVLKSEHIGP
AEGLATVGTIVTKYRPTINAVSTGDELVEPTASLRGDIRSNRMLIAAIIQ
HKQVLDLGIKATTEESLKHMDAALSDAILILSGVSMGDRDLVKCLASMGKIH
WFKIRMPKGLTFAEIVTDSKSKVLAFLPNDGSCVVCVFNLFVVPARSLSG
WSPNHPRVHARLLHPLRGDPHREHRAVIRWPDGSGRPGYVAESVQQAASSLL
SKMSALLLEVLVQMLEAGTSMQAILSDMTSPDKLPANSPSLHLPFSKSV
STDLSQVPSQNTKVAITLVSDTVSLGAGDRSGPRAISVNVSSSEKLGATVAT
AVPDDVEIKILVKSIDRSLILTLGGTGFTPRDVTBEATKSVIEKAPGLTV
MLQESLKVTFPMLSRATAGIRGSTLILINPGNPNVAVSCMEALLPALKHALKQIKGD
KREKHPRVPHAAAPVDQMDKSFRAASSGRGSCRP"
complement(11086..17771)
/gene="OSJNBa0084K01.3"
complement(join(11086..11128,11951..12019,12589..12645,
12738..13527,13604..13879,13950..14016,14110..14184,
14517..14711,14826..14975,15051..15421,15969..16061,
16145..16402,17037..17079,17667..17771))
/gene="OSJNBa0084K01.3"
/codon_start=1
/protein_id="CAE04831.2"
/db_xref="GI:38346064"
/translation="MPMQPPLAAGNWPDPHAAOAPPPSYPPPPGOGVNHWSAQ
SHGTAAGNEEDWATKAWAAKSVTEHHIOHAISSANRPNHNGYDOYQOPAG
LPAPLPPIPPQSSNDQLPQTPQTASMRDPFPQAPSSWHIAPSLQPPHFGDQSS
NYTSTFGVGNAMIQYSSQTPQTPQTASMRDPFPQAPSSWHIAPSLQPPHFGDQSS
KIAVDPSDQPLEBNSKAPDMVHRTANFNS:PAAPAAASHDHAISAPQWIPFTV
VLFPQASVPPQAQMDPVSAAFLFVAGVSSSYIPPAAFGVGNVTFAPVDANTFNV
TESKPPVPNRLREILKKKTSVSVAHSEFFHSGKSEDAEKLTKSDQADSSVS
DSAKSTDDEDEIEAARMAANOIEKVLTEVLLKVTDDLFDIATKVLNDDPS
AEPNPTRVSSVKGPGKSKPKVAKVVPKPTNVSSADHSDGSLSPKGLLGL
ASVSDDEDEGDNEDKVPVSNUSSETKAGARPGEKATDGEQGNHNRNVQCN
ASLGEDRKNDEMSHRNSAELKQELRHDTHEFFPSDAKISSQSGNHTLDEKAKH
GYSQNGKVPSPNNVFNFSNMQHGLGKSSNEEDFVKEPKAVRKRKSESSSKRYND

```

DKKSYGNIDKSGKPKKSGDRSVRHGADTREPRSRGNSKODDAKGERKDYQKDV
EKSRDADRREKDEKRSQRTKSSSHSRSRSPSARSRTKSSSHSRSVSS
DEFSNRSKRKLHSCSKSSMSPPPKSRNRVRSRSPHSKSHSRHSYPSAERKKRSR
RTPVKRR"
/sequence (20406..21580)
/gene="OSJNBa0084K01.4"
complement (join(20406..20487,20766..20841,20927..21093,
21291..21360,21443..21580))
/gene="OSJNBa0084K01.4"
/codon_start=1
/protein_id="CAE04832.1"
/db_xref="GI:32489035"
/translation="MATSSAYPPPPYLYKYDYKDPSSAPPPPPVGGYQLFGAT
YTDVPLPSDQGVQLPKSPDIDFKELRLAKREQLHLELADILVPRPOYAR
RVDSLIFKNLHLLNSLRPHQARATLHMLENQIRKQAIEDIKORREEAQKLIG
ESLLILDGNQPSLPAN"
23370..24831
/gene="OSJNBa0084K01.5"
join(23370..23556,23787..23879,23963..24044,24798..24831)
/gene="OSJNBa0084K01.5"
/codon_start=1
/protein_id="CAE04833.2"
/db_xref="GI:38346065"
/translation="MDVVKAAQLSGRTLEKVVVHPLVLLSIVDHYNRVARDTRKRVVG
VLIGTSRSGSDVNTSYAVPDEDDKDPRIWFLDHNHESFMSFKRINAKHEVVGWY
STGPKLRENDLDIHALFNHYVPQPCGDY"
27864..29507
/gene="OSJNBa0084K01.6"
join(27864..28234,29348..29507)
/gene="OSJNBa0084K01.6"
/codon_start=1
/protein_id="CAE04834.2"
/db_xref="GI:38346066"
/translation="WAFSSPSPFRRCGGGGGSPKTSASASAAAAAGNEEBEL
GTVPLQLFLRLSPSKSALQOGNSHTLAISISSPPLSGDLDRVACFSQG
WKLNSPAGRSQTPCSPKPRYELRAIQEKVRMETENGAKDVTIVEMQESK
SRDSQTLPCDESDQS"
30293..31927
/gene="OSJNBa0084K01.7"
complement (join(30293..30524,30814..30888,31535..31642,
31794..31927))
/gene="OSJNBa0084K01.7"
/codon_start=1
/protein_id="CAE04835.2"
/db_xref="GI:38346067"
/translation="WRKLPFEQKLLKKTFLDFKREKGRDAIVTORYLLVERDDYK
KYNGICLVQKLVNLIKMDPRDPYRAEMTDLKLYNMGVITKSLTCEKLSVS
SFCCRRLATVVKLAFABHLEAVTIIQQGHVRVGPETVTPAPFLVTRNMEDFITWVD
SKIKRKVOEYNGLEDLDFDAMA"
33553..35984
/gene="OSJNBa0084K01.8"
complement (35553..35984)
/gene="OSJNBa0084K01.8"
/codon_start=1
/protein_id="CAE04836.2"
/db_xref="GI:38346068"
/translation="MAKSKIRYIWMLEQTLRWRWSRAAAARAAAVPAGHVAVCVG
ASRFPVRAAHNHPFVRLROAEYFGFPAGYCGPIALPCDEGLFEHVLRLSSP
SSARFVLEDTQSCALSCCAAGDALPLRLGIATDKAVN"
51031..51031
/gene="OSJNBa0084K01.9"
complement (50570..51031)
/gene="OSJNBa0084K01.9"
/codon_start=1
/protein_id="CAE04837.2"
/db_xref="GI:38346069"
/translation="MCKVITIQSLVWLRTVRRWSRAAARPAKSKMERDGDVAPGAGH
VAVRVGGGSDARFVVRVAHLNHPFVRLROAEYFGFPAGYCGPIALPCDEGLFEH
DVLRRVSDERHDLAFPCPATCGVTKRAESSSSSRAPLQGMDSLVN"
55313..57999
/gene="OSJNBa0084K01.10"
complement (join(55313..55654,55732..56160,57637..57999))
/gene="OSJNBa0084K01.10"

```

```

/codon_start=1
/protein_id="CAE04838.2"
/db_xref="GI:38346070"
/translation="MAPVATITFLPTASNEATLPSVRDEDERPRVAYNQFSDAVPVI
SLQIGZARAEIRARVAGACEEWGIFQVVDHGVADGLVADMARLARDFFALPPEDKL
RFMSGKNGGFIYSSHLOGEAVKDWREIVTYSYFVKGRDYSRWPKFAGWRAVVEQ
YSRLMGLACKGLVLSSEAVGADTNALACVMDQKVVNFYFKPCQPDLTGLKRRH
TDEGTTITLLQDLVGGLOATRDAGTKWITVQPIPGSFVNVNLGDHAYLSNGRFPKNADH
Query Match 79.1% Score 17.4; DB 8; Length 163448;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCCGCCT 19
||| ||||| ||||| |||||
Db 27924 CCGAGCAAAACCTCCGCCT 27942
RESULT 23
BX510331 176859 bp DNA linear HTG 24-SEP-2003
LOCUS Danio rerio clone CH2111-42B22, WORKING DRAFT SEQUENCE, 3 unordered
DEFINITION
BX510331
ACCESSION BX510331
VERSION BX510331.5 GI:35209544
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 176859)
McLaren,S.
Direct Submission
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk
On Sep 24, 2003 this sequence version replaced gi:31559357.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC42B22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 175562 bases at least Q40
Consensus quality: 175799 bases at least Q30
Consensus quality: 175979 bases at least Q20
Insert size: 176659; sum-of-contigs
Quality score: 187706; 3.2% error; agarose-fp
Coverage: 8.99x in Q20 bases; sum-of-contigs Quality
coverage: 8.61x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 60644: contig of 60644 bp in length
* 60645 60744: gap of 100 bp
* 60745 128253: contig of 67509 bp in length
* 128254 128553: gap of 100 bp
* 128354 176859: contig of 48506 bp in length.
Location/Qualifiers
1..176859
/organism="Danio rerio"
/mol_type="genomic DNA"
FEATURES
Source

```



```

/db xref="taxon:7955"
/clon="CH211-42B22"
/clone_lib="CHORI-211"
1..60644
/notes=assembly_fragment:02137
fragment_chain:1
clone_end:T7
vector_side:left"
60745..128253
/notes=assembly_fragment:01319
fragment_chain:1"
128354..176859
/notes=assembly_fragment:01889
clone_end:SP6
vector_side:right"

ORIGIN
Query Match          79.18; Score 17.4; DB 2; Length 176859;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 54417 CTGGGCAAAACCTCGGCT 54435

RESULT 24
AC095761/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-9015, *** SEQUENCING IN PROGRESS ***,
9 unordered pieces.
AC095761
AC095761.4 GI:22773150
VERSION
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 325612)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,W.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunarane,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensunewa,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,X., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,

```

```

Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzos,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snares,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 325612)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 325612)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 10, 2002 this sequence version replaced gi:21722773.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDSK
Center clone name: CH230-9015
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 219952 bases at least Q40
Consensus quality: 225538 bases at least Q30
Consensus quality: 229145 bases at least Q20
Estimated insert size: 231139; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 286218: contig of 286218 bp in length
* 286219 286318: Gap of unknown length
* 293639 293688: contig of 7370 bp in length
* 293689 293788: Gap of unknown length

```

```

* 293789 294953: contig of 1165 bp in length
* 294954 295053: gap of unknown length
* 295054 296119: contig of 1066 bp in length
* 296120 296219: gap of unknown length
* 296220 297716: contig of 1496 bp in length
* 297716 297815: gap of unknown length
* 297816 299694: contig of 1879 bp in length
* 299695 299794: gap of unknown length
* 299795 301330: contig of 1535 bp in length
* 301330 301430: gap of unknown length
* 301430 302965: contig of 1536 bp in length
* 302966 303066: gap of unknown length
* 303066 325612: contig of 22547 bp in length.

```

FEATURES

```

source
  1..325612
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-9015"
    633..1430
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH344284"
    656..11330
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH344284"
    274711..276476
    /note="wgs_end-extension"
    clone_end:Sp6"

misc_feature
  79.1%; Score 17.4; DB 2; Length 325612;
  Best Local Similarity 94.7%; Pred. No. 1.2e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

misc_feature
  79.1%; Score 17.4; DB 2; Length 325612;
  Best Local Similarity 94.7%; Pred. No. 1.2e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

ORIGIN

```

Query Match 79.1%; Score 17.4; DB 2; Length 325612;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCGCCT 19
   |||||
DB 69074 CCGGGCAAAACCTCGCT 69056

```

```

RESULT 25
SEU43356
LOCUS
DEFINITION
  SEU43356 2171 bp DNA linear BCT 12-APR-2001
  Salmonella enterica isocitrate lyase (aceA) gene, partial cds,
  isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete
  cds.
ACCESSION
  U43356
VERSION
  U43356.1 GI:1151303
KEYWORDS
  Salmonella enterica
  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Salmonella.
  1 (bases 1 to 2171)
  Nelson, K., Wang, F.S. and Selander, R.K.
  Direct Submission
  Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
  208 Mueller Lab, University Park, PA 16802, USA
  2 (bases 1 to 2171)
  Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.
  Molecular genetic relationships of the salmonellae
  Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  97076912
  MEDLINE
  PUBMED
  8975610
  3 (bases 1 to 2171)
  Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.
  Size and sequence polymorphism in the isocitrate dehydrogenase
  kinase/phosphatase gene (aceK) and flanking regions in Salmonella
  enterica and Escherichia coli
  Genetics 147 (4), 1509-1520 (1997)
  JOURNAL

```

```

MEDLINE 98072475
PUBMED 9409817
FEATURES
  source
    Location/Qualifiers
      1..2171
        /organism="Salmonella enterica"
        /mol_type="genomic DNA"
        /strain="S2980"
        /sub_species="IIia"
        /db_xref="taxon:28901"
      1..213
        /gene="aceA"
      <1..213
        /gene="aceA"
        /codon_start=1
        /transl_table=11
        /product="isocitrate lyase"
        /protein_id="AAC43890.1"
        /db_xref="GI:1151304"
        /translation="LAHAYAQEGMKHYVEKVQPEFAAAKDGITFVSHQQVGVGTGYF
        DKVTIIQGGASSVTALTGSTESQF"
      307..2091
        /gene="aceK"
      307..2091
        /gene="aceK"
        /codon_start=1
        /transl_table=11
        /product="isocitrate dehydrogenase kinase/phosphatase"
        /protein_id="AAC43891.1"
        /db_xref="GI:1151305"
        /translation="NSRGHELLIAQTILQGDQYGRLEVTSGAQORFEQADHVAQV
        QAKSRHLYDHVGLVLEQLRCLTDGKNTDANFLRVKHEVTRLLPDYPRPEIAESF
        ENSVCRLPDHESLTPERLPFESSQPERFRTPPLAKDPPEHGWELLNRLISDL
        PLRLPQWKSRIIRIILAHLELIGEDALPRLCHQVANELFYRNKAALVGLKRLPDG
        TLPFLPHRTDEGLFVDTCTTAAESIVFGFARSVMVYAPLPAALVLELRLP
        GKTAELNMAICQKHAKTESYREYLCYLAESDEFIEAPGIRGMVLFVTLPGFDRV
        FKIIKDPAPQKGAHVACYQLVKEHDRVGRVADQEFENFVLDKRDIDPALMA
        LQEAPEKITDLGEHIVIRHLYERRMVLNWLVEQVEGQQLRDAIEYGNALROLAA
        ANIFGDMLEKFNFGVTRHGRVVFYDYDEICVYTNVFRDIIPARVPEDLASEPWYSV
        SPGDVPEFRHMLCADPRIGLFEEMHADLFRADYWRALQTRIKEGHVEDVYAYRR
        QRFVSRYDIDRLDXAPAPASGNVCRTA"

ORIGIN
Query Match 78.2%; Score 17.2; DB 1; Length 2171;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCGCCTAAC 22
   |||||
DB 1132 CCGGGCAAAACCAACCGCGAGC 1153

RESULT 26
SEU43357
LOCUS
DEFINITION
  SEU43357 2171 bp DNA linear BCT 12-APR-2001
  Salmonella enterica isocitrate lyase (aceA) gene, partial cds,
  isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete
  cds.
ACCESSION
  U43357
VERSION
  U43357.1 GI:1151306
KEYWORDS
  Salmonella enterica
  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Salmonella.
  1 (bases 1 to 2171)
  Nelson, K., Wang, F.S. and Selander, R.K.
  Direct Submission
  Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
  208 Mueller Lab, University Park, PA 16802, USA
  2 (bases 1 to 2171)
  Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.
  Molecular genetic relationships of the salmonellae
  Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
  JOURNAL

```

97076912		Nelson,K., Wang,F.S. and Selandar,R.K.	
8975610		Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ., 208 Mueller Lab, University Park, PA 16802, USA	
3 (bases 1 to 2171)		(bases 1 to 2467)	
Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphatase gene (acek) and flanking regions in Salmonella enterica and Escherichia coli		Molecular genetic relationships of the salmonellae	
Genetics 147 (4), 1509-1520 (1997)		Appl. Environ. Microbiol. 62 (3), 804-808 (1996)	
98072475		97076912	
9409817		8975610	
Location/Qualifiers		Location/Qualifiers	
1. .2171		1. .2467	
/organism="Salmonella enterica"		/organism="Salmonella enterica"	
/mol_type="genomic DNA"		/mol_type="genomic DNA"	
/strain="S2983"		/strain="S2978"	
/sub_species="IIa"		/sub_species="IIIB"	
/db_xref="taxon:28901"		/db_xref="taxon:28901"	
1. .213		1. .213	
/gene="aceA"		/gene="aceA"	
<1. .213		<1. .213	
/gene="aceA"		/gene="aceA"	
/codon_start=1		/codon_start=1	
/transl_table=1		/transl_table=1	
/product="isocitrate lyase"		/product="isocitrate lyase"	
/protein_id="AAC43892.1"		/protein_id="AAC43878.1"	
/db_xref="GI:1151307"		/db_xref="GI:1151286"	
/translation="LAHAYAQEGMKHYVEKVOQPEFAAAKDGYTFVSHQOEVGTGYF		/translation="LAHAYAQEGMKHYVEKVOQPEFAAAKDGYTFVSHQOEVGTGYF	
DKVTIIQGGASSVTALTGSTESQF"		DKVTIIQGGASSVTALTGSTESQF"	
307. .2091		314. .2065	
/gene="acek"		/gene="acek"	
307. .2091		314. .2065	
/gene="acek"		/gene="acek"	
/codon_start=1		/codon_start=1	
/transl_table=1		/transl_table=1	
/product="isocitrate dehydrogenase kinase/phosphatase"		/product="isocitrate dehydrogenase kinase/phosphatase"	
/protein_id="AAC43893.1"		/protein_id="AAC43879.1"	
/db_xref="GI:1151308"		/db_xref="GI:1151287"	
/translation="MSRGLLELLIAQTILQGFDAQYGRFLEVTSGAQRFEQADWHAQV		/translation="MPCRELLIAQTILQGFDAQYGRFLEVTSGAQRFEQADWHAQV	
QMSKRHLHYDHYVGLVEQLRITDGNKTDANFLKVEKHYTRLPLDPYPRFIAESF		QMSKRHLHYDHYVGLVEQLRITDGNKTDANFLKVEKHYTRLPLDPYPRFIAESF	
FNSVYCRFLDHSRSLTPELRFPSQPERRTIPRLAKDFPDHGWELLLILSDL		FNSVYCRFLDHSRSLTPELRFPSQPERRTIPRLAKDFPDHGWELLLILSDL	
PLRLPWKNSRDRIYIIAHLTEILGEDALPRCHQVANELFYRNKAALVGLKLTDPG		PLRLPWKNSRDRIYIIAHLTEILGEDALPRCHQVANELFYRNKAALVGLKLTDPG	
TLPLPIHRTDEGELFVDITCTTTAEASIVFGFARSYVMVYAPLPAALVLEWLEILP		TLPLPIHRTDEGELFVDITCTTTAEASIVFGFARSYVMVYAPLPAALVLEWLEILP	
GKTTAEALYMAICQKHAKTESYREYLCVLAESDEKFIAPGIRGMVMTLPLGFDRV		GKTTAEALYMAICQKHAKTESYREYLCVLAESDEKFIAPGIRGMVMTLPLGFDRV	
FKIIDKPAQKEMSAHVACQYLVKEDRVGRMADTQEFENFVLDKQIDPALMAL		FKIIDKPAQKEMSAHVACQYLVKEDRVGRMADTQEFENFVLDKQIDPALMAL	
LQOAPKPTIDLGEHIVIRHLTYERRMVPLNIWLEQEGQDRDAIEEYGNAIRQLAA		LQOAPKPTIDLGEHIVIRHLTYERRMVPLNIWLEQEGQDRDAIEEYGNAIRQLAA	
ANIIFPGMLFKNFGVTRHGRVVFYDYDEICYMTEVNFDPPIPPARYPEDELAEPWYSV		ANIIFPGMLFKNFGVTRHGRVVFYDYDEICYMTEVNFDPPIPPARYPEDELAEPWYSV	
SPGVFPEEFHMLCADPRIGPLPEEMHADLFRADYWRALQTRIKEGHVEDVYVYRRR		SPGVFPEEFHMLCADPRIGPLPEEMHADLFRADYWRALQTRIKEGHVEDVYVYRRR	
ORFSVRYDIDRLDKAPASGNVCRTA"		ORFGVRYGAISSTANSS"	
ORIGIN		ORIGIN	
Query Match 78.2%; Score 17.2; DB 1; Length 2171;		Query Match 78.2%; Score 17.2; DB 1; Length 2467;	
Best Local Similarity 86.4%; Pred. No. 1.7e+03;		Best Local Similarity 86.4%; Pred. No. 1.7e+03;	
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Qy 1 CCGGGCAAAACCTCGCCTAAC 22		Qy 1 CCGGGCAAAACCTCGCCTAAC 22	
Db 1132 CCGGGCAAAACACCGCGGAGC 1153		Db 1139 CCGGGCAAAACACCGCGGAGC 1160	
RESULT 27		RESULT 28	
SEU43350		SEU43351	
LOCUS		LOCUS	
DEFINITION		DEFINITION	
Salmonella enterica isocitrate lyase (aceA) gene, partial cds, isocitrate dehydrogenase kinase/phosphatase (acek) gene, complete cds.		Salmonella enterica isocitrate lyase (aceA) gene, partial cds, isocitrate dehydrogenase kinase/phosphatase (acek) gene, complete cds.	
U43350.1 GI:1151285		U43350.1 GI:1151285	
ACCESSION		ACCESSION	
VERSION		VERSION	
KEYWORDS		KEYWORDS	
SOURCE		SOURCE	
ORGANISM		ORGANISM	
REFERENCE		REFERENCE	

ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 2468)
AUTHORS Nelson, K., Wang, F.S. and Selander, R.K.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1995) Kimberly Nelson, Biology, Penn State Univ,
208 Mueller Lab, University Park, PA 16802, USA
REFERENCE 2 (bases 1 to 2468)
AUTHORS Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.
TITLE Molecular genetic relationships of the salmonellae
JOURNAL Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
MEDLINE 97076912
PUBMED 8975610
REFERENCE 3 (bases 1 to 2468)
AUTHORS Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.
TITLE Size and sequence polymorphism in the isocitrate dehydrogenase
kinase/phosphatase gene (aceK) and flanking regions in Salmonella
enterica and Escherichia coli
JOURNAL Genetics 147 (4), 1509-1520 (1997)
MEDLINE 98072475
PUBMED 9409817
FEATURES
Location/Qualifiers
1..2468
/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="S2993"
/sub_species="II"
/db_xref="taxon:28901"
1..213
/gene="aceA"
<1..213
/gene="aceA"
/codon_start=1
/transl_table=11
/product="isocitrate lyase"
/protein_id="AAC43872.1"
/db_xref="GI:1151277"
translation="LAHAYAGGGMKHYEKVQPEFAAAKDGITFVSHQEQVGTGYF
DKVTIIQGGASSYALTGTSTESQF"
315..2066
/gene="aceK"
315..2066
/gene="aceK"
/codon_start=1
/transl_table=11
/product="isocitrate dehydrogenase kinase/phosphatase"
/protein_id="AAC43873.1"
/db_xref="GI:1151278"
translation="MPRGLELLIAOTILQGDQYGRFLEVTSGAQORFEQADWHAQV
QAKSRILYDHHVGLVLEQLRCITDGKSTDAFLRVKHYTRLPLDPYPRFIAESF
FNSVYCLFDRHSITPERLIFPSQPERRTIPRLAKDFPDHGKESILMLRLSDL
PLRUPWQKGRDIRIILAHUTETLGEDALPRCHQVAVNELFYRNKAANLVGKLTTPDG
TLPELPIHRTDEGLFVDLTITAEASIVFGFARSYFWYAPLPALVWELREILP
KTTAELYMAIGCCOKHAKTESYREYLCYLAESDEKFTAEFGIRGMVNLVETLPGFDRV
PKTIKDFAPKQSAARHACVACVQLVKEHVRGMADTQBFNFVLDKQIDPALMAL
LRQAPBKITDLGQIVICHLYIERRVPLNIMLEOVGGQLRDAIEEYGNALRQLAA
ANIIFGMLFNFGVTHGRVYVDYDEICVMTVEVNFREDIPPARYBEDELASBPWYSV
SPGDVFEERFHNLCADPRIGPLFEENHADLFRADYWRALQTRIKEGHIEDVYVYRR
QRFSVRYGAISSTANSS"

ORIGIN
Query Match 78.2%; Score 17.2; DB 1; Length 2468;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
|||||
Db 1140 CCGGGCAAAACCGCGGAGC 1161
|||||

Search completed: June 20, 2004, 11:43:17
Job time : 621.376 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 126.413 Seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 cggggcaaacctccgcctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 300 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001as.*

5: Geneseq2001bs.*

6: Geneseq2002s.*

7: Geneseq2003as.*

8: Geneseq2003bs.*

9: Geneseq2003cs.*

10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	110000	4	AAI95682_32
C 2	22	100.0	110000	4	AAI95683_32
C 3	21	95.5	723	4	AAH52060
C 4	17.8	80.9	2020	4	ABL03297
C 5	17.8	80.9	2629	8	ACF35859
C 6	17.8	80.9	4020	4	ABL03296
C 7	17.4	79.1	511	6	ABQ14264
C 8	17.4	79.1	511	6	ABQ14265
C 9	17.2	78.2	647	6	ABQ35568
C 10	17.2	78.2	647	6	ABQ35569
C 11	17.2	78.2	5925	6	ABL33576
C 12	16.8	76.4	364	3	AAA43296
C 13	16.8	76.4	502	6	ABQ30690
C 14	16.8	76.4	502	6	ABQ30691
C 15	16.8	76.4	502	6	ABQ18020
C 16	16.8	76.4	502	6	ABQ18021
C 17	16.8	76.4	536	6	ABN75186
C 18	16.8	76.4	565	6	ABQ25017
C 19	16.8	76.4	565	6	ABQ25016
C 20	16.4	74.5	771	7	ACA57563
C 21	16.4	74.5	927	4	AA53792
C 22	16.4	74.5	1077	7	ABZ38654
C 23	16.4	74.5	1320	7	ACA48845

24	16.4	74.5	3832	7	ABV76014
C 25	16.4	74.5	3963	4	ABL03639
C 26	16.4	74.5	10644	4	ABL03638
C 27	16.4	74.5	56485	3	AAH81476
C 28	16.4	74.5	110000	3	AAH81489
C 29	16.4	74.5	349980	3	AAF21612
C 30	16.2	73.6	461	7	ABZ52802
C 31	16.2	73.6	753	3	AAH69022
C 32	16.2	73.6	812	4	AAH00347
C 33	16.2	73.6	819	4	AAH00348
C 34	16.2	73.6	830	4	AAH00346
C 35	16.2	73.6	856	6	ABQ19276
C 36	16.2	73.6	856	6	ABQ19277
C 37	16.2	73.6	1911	3	AAH88709
C 38	16.2	73.6	1911	3	AAA95482
C 39	16.2	73.6	3501	9	ADD45068
C 40	16.2	73.6	4343	5	AA594174
C 41	16.2	73.6	4995	5	AA582543
C 42	16.2	73.6	6022	4	AA545374
C 43	16.2	73.6	6022	4	AA546397
C 44	16.2	73.6	6022	6	ABL33072
C 45	16.2	73.6	6022	6	ABX28211
C 46	16.2	73.6	9888	6	ABL33241
C 47	16.2	73.6	16668	3	AA569013
C 48	15.8	71.8	20	6	AB197066
C 49	15.8	71.8	24	6	AB197073
C 50	15.8	71.8	24	6	AB197072
C 51	15.8	71.8	346	6	AAH34474
C 52	15.8	71.8	828	6	ABQ44419
C 53	15.8	71.8	828	6	ABQ44418
C 54	15.8	71.8	829	6	ABQ52905
C 55	15.8	71.8	829	6	ABQ52904
C 56	15.8	71.8	954	2	AAV19862
C 57	15.8	71.8	1053	6	ABQ42128
C 58	15.8	71.8	1053	6	ABQ42129
C 59	15.8	71.8	1101	6	AAK99715
C 60	15.8	71.8	1224	4	ABL21641
C 61	15.8	71.8	1280	2	AAQ05610
C 62	15.8	71.8	1281	1	AAH50059
C 63	15.8	71.8	1347	2	AAV19864
C 64	15.8	71.8	1375	9	ACC00840
C 65	15.8	71.8	1375	9	ADC23681
C 66	15.8	71.8	1597	2	AAV19863
C 67	15.8	71.8	2333	2	AAV19861
C 68	15.8	71.8	2307	7	ACA24522
C 69	15.8	71.8	5027	2	AAV19860
C 70	15.8	71.8	11262	4	ABL29038
C 71	15.8	71.8	22507	6	AB578886
C 72	15.8	71.8	28344	4	ABL21640
C 73	15.6	70.9	117	6	ABL85045
C 74	15.6	70.9	403	7	ABX38154
C 75	15.6	70.9	558	7	ACA52390
C 76	15.6	70.9	662	6	ABQ16296
C 77	15.6	70.9	662	6	ABQ16297
C 78	15.6	70.9	815	6	ABQ15362
C 79	15.6	70.9	815	6	ABQ15363
C 80	15.6	70.9	876	8	ACA30165
C 81	15.6	70.9	882	7	ACF70675
C 82	15.6	70.9	957	7	ACA38950
C 83	15.6	70.9	969	6	ABQ50158
C 84	15.6	70.9	969	6	ABQ50159
C 85	15.6	70.9	1041	3	AAE55804
C 86	15.6	70.9	1041	3	ADE10223
C 87	15.6	70.9	1182	7	ACA36111
C 88	15.6	70.9	1231	4	ACA32670
C 89	15.6	70.9	1884	3	AAAC1212
C 90	15.6	70.9	2098	1	AAH91208
C 91	15.6	70.9	2994	2	AAQ10500
C 92	15.6	70.9	3344	5	AA585570
C 93	15.6	70.9	5059	3	AAZ36227
C 94	15.6	70.9	5663	6	ABL32697
C 95	15.6	70.9	5663	6	ABL32697
C 96	15.6	70.9	7029	6	ABL92223
C 97	15.6	70.9	7029	6	AA517119

ABV76014	Human EGF
Abi03639	Drosophila
Abi03638	Drosophila
AA81476	N. mening
Continuation (5 of	
Abz21612	Neisseria
Abz52802	Aspergillus
AAH00347	Providencia
AAH00348	Providencia
Abq19276	Oligonucle
Abq19277	Oligonucle
AAH88709	E. coli F
AA95482	E. coli e
ADD45068	Rat gene
AA594174	DNA encod
AA582543	DNA encod
AA545374	Chemical
AA546397	Tumour su
ABL33072	Human imm
ABX28211	DNA trans
ABL33241	Human imm
AA569013	Bacteriop
AB197066	Capture o
AB197073	Capture o
AB197072	Capture o
AAH34474	Human col
ABQ44419	Oligonucle
ABQ44418	Oligonucle
ABQ52905	Oligonucle
ABQ52904	Oligonucle
AAV19862	Gallus do
ABQ42128	Oligonucle
ABQ42129	Oligonucle
AAK99715	DNA of ni
ABL21641	Drosophila
AAQ05610	nifH gene
AAH50059	Sequence
AAV19864	Gallus do
ACC00840	Glycine m
ADC23681	CDNA enco
AAV19863	Gallus do
AAV19861	Gallus do
ACA24522	Prokaryot
AAV19860	Gallus do
ABL29038	Drosophila
AB578886	E. coli C
ABL21640	Drosophila
ABL85045	Human ova
ABX38154	Bovine ES
ACA52390	Prokaryot
ABQ16296	Oligonucle
ABQ16297	Oligonucle
ABQ15362	Oligonucle
ABQ15363	Oligonucle
ACA30165	DNA encod
ACF70675	Phototrab
ACA38950	Prokaryot
ABQ50158	Oligonucle
ABQ50159	Oligonucle
AAE55804	S. lavend
ADE10223	S. lavend
ACA36111	Prokaryot
ACA32670	Human cdn
AAAC1212	Arabidops
AAH91208	Acetyl-Co
AAQ10500	Acetacet
AA585570	DNA encod
AAZ36227	DNA encod
ABL32697	Human imm
ABL32697	Human imm
ABL92223	Chemical
AA517119	Ketogulon

C 97	15.6	70.9	8115	6	ABA97958	AbA97958	Lactic ac	C 170	15.2	69.1	2452	2	AAV61938	Mouse CFL
C 98	15.6	70.9	8115	6	ABA97959	AbA97959	Lactic ac	C 171	15.2	69.1	2494	9	AD782290	Endometri
C 99	15.6	70.9	10183	4	AA546752	Aax46752	Tumour su	C 172	15.2	69.1	2896	6	AA777712	Murine pl
C 100	15.6	70.9	19483	2	AAX20541	Aax20541	Polynucle	C 173	15.2	69.1	2770	6	AB152325	Mouse FLI
C 101	15.6	70.9	19513	6	AA561246	AA561246	Human gen	C 174	15.2	69.1	2770	7	ABV72474	Prokaryot
C 102	15.6	70.9	19513	6	ABL49340	ABL49340	Human pol	C 175	15.2	69.1	3741	7	ACA50830	Prokaryot
C 103	15.6	70.9	31940	4	AF28526	Aaf28526	Genomic f	C 176	15.2	69.1	4395	7	ACA48474	Prokaryot
C 104	15.6	70.9	43411	6	ABQ88169	Abq88169	Human ost	C 177	15.2	69.1	5857	6	AB132528	Human imm
C 105	15.6	70.9	53500	3	AA55842	AAc55842	Complete	C 178	15.2	69.1	5857	6	AB132528	Human imm
C 106	15.6	70.9	53500	3	AA55842	AAc55842	Complete	C 179	15.2	69.1	5857	6	AB132528	Human imm
C 107	15.6	70.9	96593	8	ADA02601	AdA02601	S. lavend	C 180	15.2	69.1	33962	4	AB16740	Drosophil
C 108	15.6	70.9	96593	8	ADA02601	AdA02601	S. lavend	C 181	15.2	69.1	33962	4	AB16740	Drosophil
C 109	15.6	70.9	96593	8	ADA02601	AdA02601	S. lavend	C 182	15.2	69.1	33962	4	AB16740	Drosophil
C 110	15.6	70.9	110000	7	ACF67367	ACF67367	Human RUN	C 183	15.2	69.1	110000	3	AA81458	N. mening
C 111	15.6	70.9	110000	7	ACF67367	ACF67367	Human RUN	C 184	15.2	69.1	110000	3	AA81458	N. mening
C 112	15.6	70.9	34980	5	APF66431	Aaf66431	Pyrococu	C 185	15.2	69.1	110000	3	AA81458	N. mening
C 113	15.4	70.0	408	5	APF66431	Aaf66431	Pyrococu	C 186	15.2	69.1	110000	3	AA81458	N. mening
C 114	15.4	70.0	514	6	AA561502	Aas61502	Lung smal	C 187	15.2	69.1	110000	4	AA199683	36
C 115	15.4	70.0	616	4	AA546863	Aas46863	Human G p	C 188	15.2	69.1	172325	3	AA21613	Neisseria
C 116	15.4	70.0	616	4	AA546863	Aas46863	Human G p	C 189	15.2	69.1	172325	3	AA21613	Neisseria
C 117	15.4	70.0	638	6	ABQ35553	Abq35553	Oligonuc	C 190	15.2	69.1	34980	3	AA21613	Neisseria
C 118	15.4	70.0	638	6	ABQ35553	Abq35553	Oligonuc	C 191	15.2	69.1	34980	3	AA21613	Neisseria
C 119	15.4	70.0	662	6	ABQ49882	Abq49882	Oligonuc	C 192	15.2	69.1	34980	3	AA21613	Neisseria
C 120	15.4	70.0	730	6	ABQ51576	Abq51576	Oligonuc	C 193	15.2	69.1	34980	3	AA21613	Neisseria
C 121	15.4	70.0	730	6	ABQ51576	Abq51576	Oligonuc	C 194	15.2	69.1	34980	3	AA21613	Neisseria
C 122	15.4	70.0	766	6	ABQ15299	Abq15299	Oligonuc	C 195	15.2	69.1	34980	3	AA21613	Neisseria
C 123	15.4	70.0	766	6	ABQ15299	Abq15299	Oligonuc	C 196	15.2	69.1	34980	3	AA21613	Neisseria
C 124	15.4	70.0	1173	6	ABQ14876	Abq14876	Oligonuc	C 197	15.2	69.1	34980	3	AA21613	Neisseria
C 125	15.4	70.0	1173	6	ABQ14876	Abq14876	Oligonuc	C 198	15.2	69.1	34980	3	AA21613	Neisseria
C 126	15.4	70.0	1233	6	ABK75674	Abk75674	Bacillus	C 199	15.2	69.1	34980	3	AA21613	Neisseria
C 127	15.4	70.0	1314	6	ABQ32840	Abq32840	Oligonuc	C 200	15.2	69.1	34980	3	AA21613	Neisseria
C 128	15.4	70.0	1314	6	ABQ32840	Abq32840	Oligonuc	C 201	15.2	69.1	34980	3	AA21613	Neisseria
C 129	15.4	70.0	1401	4	AA522827	Aas22827	E. coli D	C 202	15.2	69.1	34980	3	AA21613	Neisseria
C 130	15.4	70.0	1401	4	AA522827	Aas22827	E. coli D	C 203	15.2	69.1	34980	3	AA21613	Neisseria
C 131	15.4	70.0	1556	5	AA522827	Aas22827	E. coli D	C 204	15.2	69.1	34980	3	AA21613	Neisseria
C 132	15.4	70.0	1556	5	AA522827	Aas22827	E. coli D	C 205	15.2	69.1	34980	3	AA21613	Neisseria
C 133	15.4	70.0	8201	6	AB133864	Ab133864	Human imm	C 206	15.2	69.1	34980	3	AA21613	Neisseria
C 134	15.4	70.0	8201	6	AB133864	Ab133864	Human imm	C 207	15.2	69.1	34980	3	AA21613	Neisseria
C 135	15.4	70.0	12460	6	AB132068	Ab132068	Human imm	C 208	15.2	69.1	34980	3	AA21613	Neisseria
C 136	15.4	70.0	31248	8	ACC58250	Acc58250	Coumermyc	C 209	15.2	69.1	34980	3	AA21613	Neisseria
C 137	15.4	70.0	35359	8	ACC58250	Acc58250	Coumermyc	C 210	15.2	69.1	34980	3	AA21613	Neisseria
C 138	15.2	69.1	366	5	ABV27608	Abv27608	Human pro	C 211	15.2	69.1	34980	3	AA21613	Neisseria
C 139	15.2	69.1	366	5	ABV27608	Abv27608	Human pro	C 212	15.2	69.1	34980	3	AA21613	Neisseria
C 140	15.2	69.1	376	5	ABV11940	Abv11940	Human pro	C 213	15.2	69.1	34980	3	AA21613	Neisseria
C 141	15.2	69.1	392	5	ABV02771	Abv02771	Human pro	C 214	15.2	69.1	34980	3	AA21613	Neisseria
C 142	15.2	69.1	392	5	ABV02771	Abv02771	Human pro	C 215	15.2	69.1	34980	3	AA21613	Neisseria
C 143	15.2	69.1	424	5	ABV33085	Abv33085	Human pro	C 216	15.2	69.1	34980	3	AA21613	Neisseria
C 144	15.2	69.1	424	5	ABV33085	Abv33085	Human pro	C 217	15.2	69.1	34980	3	AA21613	Neisseria
C 145	15.2	69.1	512	6	ABQ24498	Abq24498	Oligonuc	C 218	15.2	69.1	34980	3	AA21613	Neisseria
C 146	15.2	69.1	512	6	ABQ24498	Abq24498	Oligonuc	C 219	15.2	69.1	34980	3	AA21613	Neisseria
C 147	15.2	69.1	531	6	ABQ31476	Abq31476	Oligonuc	C 220	15.2	69.1	34980	3	AA21613	Neisseria
C 148	15.2	69.1	531	6	ABQ31476	Abq31476	Oligonuc	C 221	15.2	69.1	34980	3	AA21613	Neisseria
C 149	15.2	69.1	534	9	ADC92247	Adc92247	E. faeciu	C 222	15.2	69.1	34980	3	AA21613	Neisseria
C 150	15.2	69.1	617	9	ADC76257	Adc76257	DNA hemol	C 223	15.2	69.1	34980	3	AA21613	Neisseria
C 151	15.2	69.1	645	9	ADC76257	Adc76257	DNA hemol	C 224	15.2	69.1	34980	3	AA21613	Neisseria
C 152	15.2	69.1	663	9	ADC76257	Adc76257	DNA hemol	C 225	15.2	69.1	34980	3	AA21613	Neisseria
C 153	15.2	69.1	682	4	AA115702	Aal15702	Human bre	C 226	15.2	69.1	34980	3	AA21613	Neisseria
C 154	15.2	69.1	889	4	AA194216	Aal94216	Human neu	C 227	15.2	69.1	34980	3	AA21613	Neisseria
C 155	15.2	69.1	889	7	ABZ32159	Abz32159	Human neu	C 228	15.2	69.1	34980	3	AA21613	Neisseria
C 156	15.2	69.1	1032	7	ABZ40136	Abz40136	N. gonorr	C 229	15.2	69.1	34980	3	AA21613	Neisseria
C 157	15.2	69.1	1044	6	ABQ19587	Abq19587	Oligonuc	C 230	15.2	69.1	34980	3	AA21613	Neisseria
C 158	15.2	69.1	1044	6	ABQ19587	Abq19587	Oligonuc	C 231	15.2	69.1	34980	3	AA21613	Neisseria
C 159	15.2	69.1	1050	7	ABZ40137	Abz40137	N. gonorr	C 232	15.2	69.1	34980	3	AA21613	Neisseria
C 160	15.2	69.1	1230	7	ABZ40137	Abz40137	N. gonorr	C 233	15.2	69.1	34980	3	AA21613	Neisseria
C 161	15.2	69.1	1230	7	ABZ40137	Abz40137	N. gonorr	C 234	15.2	69.1	34980	3	AA21613	Neisseria
C 162	15.2	69.1	1254	7	ABZ39897	Abz39897	N. gonorr	C 235	15.2	69.1	34980	3	AA21613	Neisseria
C 163	15.2	69.1	1281	7	ABZ39897	Abz39897	N. gonorr	C 236	15.2	69.1	34980	3	AA21613	Neisseria
C 164	15.2	69.1	1428	7	ABZ39897	Abz39897	N. gonorr	C 237	15.2	69.1	34980	3	AA21613	Neisseria
C 165	15.2	69.1	1573	2	AA233639	Aaz33639	Human bre	C 238	15.2	69.1	34980	3	AA21613	Neisseria
C 166	15.2	69.1	1881	8	ABZ69322	Abz69322	P aerigin	C 239	15.2	69.1	34980	3	AA21613	Neisseria
C 167	15.2	69.1	2000	7	ADA73033	AdA73033	Rice gene	C 240	15.2	69.1	34980	3	AA21613	Neisseria
C 168	15.2	69.1	2153	4	ABL24890	AbL24890	Drosophil	C 241	15.2	69.1	34980	3	AA21613	Neisseria
C 169	15.2	69.1	2413	6	ABL52342	AbL52342	Mouse FLI	C 242	15.2	69.1	34980	3	AA21613	Neisseria

243	14.8	67.3	902	6	ABQ46431
c 244	14.8	67.3	902	6	ABQ46430
245	14.8	67.3	918	5	AH66019
c 246	14.8	67.3	979	6	ABQ50762
247	14.8	67.3	979	6	ABQ50763
c 248	14.8	67.3	1008	7	ACF69175
249	14.8	67.3	1033	6	ABQ33359
c 250	14.8	67.3	1033	6	ABQ33358
251	14.8	67.3	1038	8	ADA29428
c 252	14.8	67.3	1047	4	AAF67916
253	14.8	67.3	1047	4	AAF71619
254	14.8	67.3	1077	6	AS97197
255	14.8	67.3	1106	5	AS93396
c 256	14.8	67.3	1143	6	AAK97114
257	14.8	67.3	1148	6	ABQ50999
c 258	14.8	67.3	1148	6	ABQ50998
259	14.8	67.3	1152	7	ACA41112
c 260	14.8	67.3	1185	7	ABZ40256
261	14.8	67.3	1185	7	ABZ41367
c 262	14.8	67.3	1218	4	AS55067
263	14.8	67.3	1243	7	ACC45235
c 264	14.8	67.3	1275	5	AH52209
265	14.8	67.3	1311	7	ACA51494
c 266	14.8	67.3	1314	4	ABL12985
267	14.8	67.3	1341	2	AAQ90492
c 268	14.8	67.3	1359	7	ACA49129
269	14.8	67.3	1371	6	AS97262
270	14.8	67.3	1374	7	ACA41912
c 271	14.8	67.3	1401	4	AS56282
272	14.8	67.3	1401	7	ACA51954
c 273	14.8	67.3	1438	4	ABL02543
274	14.8	67.3	1449	7	ACA24419
c 275	14.8	67.3	1473	7	ABZ40363
276	14.8	67.3	1483	5	AH43355
c 277	14.8	67.3	1506	7	ACF71932
278	14.8	67.3	1531	4	AH21843
c 279	14.8	67.3	1548	4	ABL30045
280	14.8	67.3	1646	4	ABL16835
c 281	14.8	67.3	1660	4	AAI58832
282	14.8	67.3	1660	8	ADB48813
c 283	14.8	67.3	1756	7	ACA35508
284	14.8	67.3	2017	4	AAI60618
c 285	14.8	67.3	2028	6	ABK76947
286	14.8	67.3	2050	9	ADB53454
c 287	14.8	67.3	2140	2	AAV42979
288	14.8	67.3	2140	2	AZ96289
c 289	14.8	67.3	2768	4	ABL18106
290	14.8	67.3	3490	4	ABL12984
c 291	14.8	67.3	3566	4	ABL02542
292	14.8	67.3	3633	2	AAQ75818
c 293	14.8	67.3	3714	4	ABL30044
294	14.8	67.3	3945	4	ABL16834
c 295	14.8	67.3	3987	3	AA81772
296	14.8	67.3	4539	4	ABL03472
c 297	14.8	67.3	5034	3	AA81492
298	14.8	67.3	5068	5	AAF32508
c 299	14.8	67.3	5080	4	ABL03231
300	14.8	67.3	5230	7	AD49464

ALIGNMENTS

RESULT 1
AAI99682_32/c
Continuation (33 of 45) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000

WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	451529

Query Match 100.0%; Score 22; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAAC 22

DB 39846 CCGGGCAAAACCTCCGCTAAC 39825

RESULT 2

AAI99682_32/c
Continuation (33 of 44) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End

WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000

Wp AAI99683 18 1800001 1910000
 Wp AAI99683 19 1900001 2010000
 Wp AAI99683 20 2000001 2110000
 Wp AAI99683 21 2100001 2210000
 Wp AAI99683 22 2200001 2310000
 Wp AAI99683 23 2300001 2410000
 Wp AAI99683 24 2400001 2510000
 Wp AAI99683 25 2500001 2610000
 Wp AAI99683 26 2600001 2710000
 Wp AAI99683 27 2700001 2810000
 Wp AAI99683 28 2800001 2910000
 Wp AAI99683 29 2900001 3010000
 Wp AAI99683 30 3000001 3110000
 Wp AAI99683 31 3100001 3210000
 Wp AAI99683 32 3200001 3310000
 Wp AAI99683 33 3300001 3410000
 Wp AAI99683 34 3400001 3510000
 Wp AAI99683 35 3500001 3610000
 Wp AAI99683 36 3600001 3710000
 Wp AAI99683 37 3700001 3810000
 Wp AAI99683 38 3800001 3910000
 Wp AAI99683 39 3900001 4010000
 Wp AAI99683 40 4000001 4110000
 Wp AAI99683 41 4100001 4210000
 Wp AAI99683 42 4200001 4310000
 Wp AAI99683 43 4300001 4403765

Query Match 100.0%; Score 22; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCTAAC 22
 Db 34122 CCGGGCAAAACCTCCGCTAAC 34101

RESULT 3
 AAH52060
 ID AAH52060 standard; DNA; 723 BP.

AC AAH52060;
 DT 04-SEP-2001 (first entry)
 DE Mycobacterium tuberculosis potential drug target gene SEQ ID 114.
 XX Drug target; growth; organism viability; characterisation; ds.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000WO-US031152.

PR 12-NOV-1999; 99US-0165086P.

PR 12-NOV-1999; 99US-0165124P.

PR 01-FEB-2000; 2000US-0179531P.

XX (REGC) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI; 2001-329193/34.

DR P-PSDB; AAG81209.

PT Identifying nucleotide or polypeptide sequence for use as drug target,
 PT involves providing algorithm that analyzes a functional relationship
 PT between nucleotide or polypeptide sequences, and comparing the sequences.

PS Disclosure; Page 128; 207pp; English.

XX

CC This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism

XX Sequence 723 BP; 137 A; 224 C; 240 G; 122 T; 0 U; 0 Other;
 Query Match 95.5%; Score 21; DB 4; Length 723;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCTAAC 21
 Db 703 CCGGGCAAAACCTCCGCTAAC 723

RESULT 4

ABL03297/c

ID ABL03297 standard; cDNA; 2020 BP.

XX ABL03297;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 4373.

XX Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB59194.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signaling and cell-cell
 XX interactions.

PS Claim 1; SEQ ID NO 4373; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 2020 BP; 500 A; 516 C; 589 G; 415 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 4; Length 2020;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 1986 CCGGGCATAACCTCCGCTGA 1966

RESULT 5
ACF35859/c
ID ACF35859 standard; cDNA; 2629 BP.
XX AC ACF35859;
XX DT 06-NOV-2003 (first entry)
XX DE D. melanogaster sphingosine kinase (SK)1 cDNA.
XX KW sphingolipid; cytosolic; gene therapy; cancer; sphingolipidase; SPL;
XX KW sphingosine-1-phosphate lyase; sphingosine kinase; SK1; gene; ss.
XX OS Drosophila melanogaster.
XX XX
XX Key Location/Qualifiers
XX FH 235..2160
XX CDS /*tag= a
XX FT /*product= "SK"
XX FT /*note= "sphingosine kinase (ABR82392)"
XX FT 688..2160
XX FT /*tag= b
XX FT /*note= "SK1 fragment (ABR82389)"
XX PN WO2003062390-A2.
XX PD 31-JUL-2003.
XX PF 17-JAN-2003; 2003WO-US001739.
XX PR 17-JAN-2002; 2002US-00053510.
XX PR 17-JAN-2002; 2002US-0349582P.
XX PA (CHIL-) CHILDRENS HOSPITAL & RES CENT AT OAKLAND.
XX PI Saba JD, Fyrt H;
XX DR WPI; 2003-663394/62.
XX DR P-PSDB; ABR82389, ABR82392.
XX PT Identifying an agent that modulates sphingolipid metabolism for treating
XX PT e.g., breast cancer by culturing a homozygous null mutant Drosophila
XX PT melanogaster in the absence and presence of a candidate agent.
XX PS Claim 4; Page 127-128; 93pp; English.
XX CC The invention relates to identifying an agent that modulates sphingolipid
XX CC metabolism. The method is useful for identifying an agent that modulates
XX CC sphingolipid metabolism for preparing a composition for treating a
XX CC disease associated with altered sphingolipid metabolism in a patient,
XX CC e.g., colon cancer, breast cancer, uterine cancer, stomach cancer,
XX CC ovarian cancer, lung cancer, kidney cancer, adenocarcinoma of the rectum
XX CC or hereditary sensory neuropathy type 1 or any one of the
XX CC sphingolipidases. The present sequence represents a D. melanogaster
XX CC sphingosine kinase (SK)1 cDNA (GI:21429173)
XX SQ Sequence 2629 BP; 721 A; 625 C; 679 G; 604 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 8; Length 2629;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 2126 CCGGGCATAACCTCCGCTGA 2106

RESULT 6
ABL03296/c
ID ABL03296 standard; cDNA; 4020 BP.
XX AC ABL03296;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 4370.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABR59193.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 4370; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR59737-
XX CC ABR72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 4020 BP; 1245 A; 857 C; 893 G; 1025 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 4; Length 4020;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 2986 CCGGGCATAACCTCCGCTGA 2966

RESULT 7
ABQ14264/c
ID ABQ14264 standard; DNA; 511 BP.
XX AC ABQ14264;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 855.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PP 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX PA (EPIC-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 511 BP; 78 A; 48 C; 185 G; 200 T; 0 U; 0 Other;
 Query Match 79.1%; Score 17.4; DB 6; Length 511;
 Best Local Similarity 94.7%; Pred. No. 1.4e-02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 GGCAAAACCTCGGCTAAC 22
 DB 25 GACAAACCTCGGCTAAC 7
 RESULT 8
 ABQ14265
 ID ABQ14265 standard; DNA; 511 BP.
 XX AC ABQ14265;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 856.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW SNP; cell differentiation; ds.

XX KW Gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PP 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX PA (EPIC-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 511 BP; 200 A; 185 C; 48 G; 78 T; 0 U; 0 Other;
 Query Match 79.1%; Score 17.4; DB 6; Length 511;
 Best Local Similarity 94.7%; Pred. No. 1.4e-02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 GGCAAAACCTCGGCTAAC 22
 DB 487 GACAAACCTCGGCTAAC 505
 RESULT 9
 ABQ35568/C
 ID ABQ35568 standard; DNA; 647 BP.
 XX AC ABQ35568;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22159.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX

PN W0200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 1549; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 5925 BP; 1510 A; 105 C; 1246 G; 3064 T; 0 U; 0 Other;
 Query Match 78.2%; Score 17.2; DB 6; Length 5925;
 Best Local Similarity 86.4%; Pred. No. 2.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 CGGGCAAAACCTCGCCTAAC 22
 DB 4911 CGGGCAAAACCTCGCCTAAC 4890
 RESULT 12
 AAA43296/c
 ID AAA43296 standard; cDNA; 364 BP.
 XX
 AC AAA43296;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Xenopus secreted expressed sequence tag SEQ ID NO:2036.
 XX
 KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
 KW antitumor; osteopathic; neuroprotective; neurotropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
 KW autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
 KW infection; depression; psoriasis; ss.
 XX
 OS Xenopus sp.
 XX
 PN W0200021990-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US024205.

XX 15-OCT-1998; 98US-0104435P.
 PR (GEM) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 XX
 DR WPI; 2000-317937/27.
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders.
 XX
 PS Claim 1; Page 585; 618pp; English.
 XX
 CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat tissue
 CC sources. The SESTs can have a range of activities depending on the
 CC tissues they were isolated from. The activities include: chemotactic;
 CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
 CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
 CC antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor;
 CC osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic;
 CC cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be
 CC used for gene therapy and in vaccines. The SESTs are useful as probes for
 CC the identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions,
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention
 XX
 SQ Sequence 364 BP; 111 A; 70 C; 100 G; 83 T; 0 U; 0 Other;
 Query Match 76.4%; Score 16.8; DB 3; Length 364;
 Best Local Similarity 90.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 CGGGCAAAACCTCGCCTAA 21
 DB 176 CGGTCAAAACCTCTGCCTAA 157
 RESULT 13
 ABQ30690/c
 ID ABQ30690 standard; DNA; 502 BP.
 XX
 AC ABQ30690;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17281.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.

01-SEP-2000; 2000DE-01043826.
05-SEP-2000; 2000DE-01044543.
(EPiG-) EPIGENOMICS AG.
Olek A, Pienbrock C, Berlin K, Guetig D;
WPI; 2002-371829/40.
Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.
Claim 12; 56pp + Sequence Listing; 56pp; German.
This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention
Sequence 502 BP; 105 A; 47 C; 139 G; 211 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 502;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY 2 CGGCAAAACCTCCGCTAA 21
DB 216 CGCGTAAACCTCCGCTAA 197
RESULT 14
ABQ30691
ID ABQ30691 standard; DNA; 502 BP.
XX AC
XX ABQ30691;
12-JUL-2002 (first entry)
XX
XX
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17282.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200218632-A2.
XX
XX
PD 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX

PA	(EPIG-) EPIGENOMICS AG.
XX	Olek A, Piepenbrock C, Berlin K, Guetig D;
XX	WPI; 2002-371829/40.
XX	Determining the degree of cytosine methylation in genomic DNA, useful for
PT	diagnosis and prognosis, comprises selective hybridization of amplicons
PT	from chemically treated DNA.
XX	Claim 12; 56pp + Sequence Listing; 56pp; German.
XX	This invention describes a novel method for determining the degree of
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one member,
CC	of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC	degree of hybridisation to both classes is determined from the label on
CC	the amplicon. From the ratio of labels hybridised to the two classes of
CC	oligomers, the degree of methylation is calculated. The method is used:
CC	(i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC	and of a wide range of diseases, e.g. cancer, disorders of the central
CC	nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC	particularly by detecting mutations or single nucleotide polymorphisms
CC	(SNP's); and (ii) for differentiation of cell or tissue types and for
CC	investigating cell differentiation. The method allows the methylation
CC	status of many C residues to be determined simultaneously. ABQ13410-
CC	ABQ54121 represent genomic DNA sequences used to illustrate the method
CC	for determining the degree of cytosine methylation described in the
CC	disclosure of the invention
XX	
SQ	Sequence 502 BP; 211 A; 139 C; 47 G; 105 T; 0 U; 0 Other;
	Query Match 76.4%; Score 16.8; DB 6; Length 502;
	Best Local Similarity 90.0%; Pred. No. 2.8e+02;
	Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps
QY	2 CGGGCAAAACCTCCGCCTAA 21
DB	287 CGCGTAAACCTCCGCCTAA 306
RESULT 15	
ID	ABQ18020/c
IX	ABQ18020 standard; DNA; 502 BP.
XX	AC ABQ18020;
XX	
DT	12-JUL-2002 (first entry)
XX	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 4611.
XX	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
XX	SNP; cell differentiation; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200218632-A2.
XX	
PD	07-MAR-2002.
XX	
PF	01-SEP-2001; 2001WO-EPO10074.
XX	
PR	01-SEP-2000; 2000DE-01043826.
PR	05-SEP-2000; 2000DE-01044543.
XX	
FA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.
 XX
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 502 BP; 96 A; 48 C; 139 G; 219 T; 0 U; 0 Other;
 XX
 Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCCGCTAA 21
 Db 308 CGCGTAAACCTCCGCTAA 289
 RESULT 16
 ABQ18021
 ID ABQ18021 standard; DNA; 502 BP.
 XX
 AC ABQ18021;
 XX
 XX
 DT 12-JUL-2002 (first entry)
 DE
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4612.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200218632-A2.
 PN
 XX 07-MAR-2002.
 PD
 XX
 XX 01-SEP-2001; 2001WO-EP010074.
 PF
 XX 01-SEP-2000; 2000DE-01043826.
 FR
 XX 05-SEP-2000; 2000DE-0104543.
 FR
 XX (EFIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI
 XX WPI; 2002-371829/40.
 DR
 XX

PT
 PT
 PT
 XX
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 502 BP; 219 A; 139 C; 48 G; 96 T; 0 U; 0 Other;
 XX
 Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCCGCTAA 21
 Db 195 CGCGTAAACCTCCGCTAA 214
 RESULT 17
 ABN75186/c
 ID ABN75186 standard; cDNA; 536 BP.
 XX
 AC ABN75186;
 XX
 XX
 DT 08-JUL-2002 (first entry)
 DE
 DE Human ORF133 cDNA, SEQ ID NO:265.
 XX
 XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiatic; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200190366-A2.
 PN
 XX 29-NOV-2001.
 PD
 XX 24-MAY-2001; 2001WO-US017076.
 PF
 XX 24-MAY-2000; 2000US-0206690P.
 FR
 XX (CURA-) CURAGEN CORP.
 PA

XX PI Leach MD, Shinkets RA;
XX WPI; 2002-106200/14.
DR P-PSDB; ABP31160.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX Claim 1; Page 335; 2508pp; English.
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (Open reading frame) 1-4534, and sequences ABN75054-
CC ABN7587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, anti-inflammatory activity, tumour inhibition activity,
CC and anti-infective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as epilepsy and Alzheimer's disease,
CC neurological disorders such as immune system disorders, disorders related to
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX SQ Sequence 536 BP; 109 A; 138 C; 142 G; 147 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 536;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCTAAC 22
DB 196 GGGCAAAAGCTCGGCTAAC 177
RESULT 18
ABQ25017
ID ABQ25017 standard; DNA; 565 BP.
XX AC ABQ25017;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 11608.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.

XX OS Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP010074.
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used;
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNPs); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ4121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX SQ Sequence 565 BP; 249 A; 158 C; 50 G; 108 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 565;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CGGGCAAAACCTCCGCTAAC 21
DB 258 CGGGTAAACCTCCGCTAAC 277
RESULT 19
ABQ25016/c
ID ABQ25016 standard; DNA; 565 BP.
XX AC ABQ25016;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 11607.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.

PN WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EF010074.
XX 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 565 BP; 108 A; 50 C; 158 G; 249 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 565;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CGGCGCAAAACCTCGCCTAA 21
DB 308 CGCGTAAACCTCGCCTAA 289
RESULT 20
ACAS7563
ID ACAS7563 standard; cDNA; 771 BP.
XX
AC ACAS7563;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human adipocyte Selected Interacting domain, SID, cDNA #650.
XX
KW Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS Homo sapiens.
XX
PN WO200286122-A2.
XX
PD 31-OCT-2002.

PF 14-MAR-2002; 2002WO-EF003768.
XX
PR 14-MAR-2001; 2001US-0275734P.
XX
PA (HYBR-) HYBRIGENICS.
XX
PI Legrain P, Daviet L;
XX
DR WPI; 2003-103412/09.
DR P-PSDB; ABU71019.
XX
XX New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes.
XX
PS Claim 7; Page 326; 382pp; English.
XX
CC The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are useful for
CC preventing or treating metabolic disorders such as obesity or diabetes.
CC The polynucleotides are useful as probes or primers. The complex is
CC particularly useful for identifying selected interacting domains (SID
CC (RTM)) for screening drugs that modulate the protein interaction, thus
CC exhibiting the therapeutic effect. The present sequence encodes a SID
CC (prey) protein of the invention
XX
SQ Sequence 771 BP; 232 A; 173 C; 188 G; 178 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 7; Length 771;
Best Local Similarity 94.4%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 CGGCGCAAAACCTCGCCT 19
DB 404 CGGCGCAAAACCTCCACCT 421
RESULT 21
AAS53792/c
ID AAS53792 standard; DNA; 927 BP.
XX
AC AAS53792;
XX
DT 13-FEB-2002 (first entry)
XX
DE Helicobacter pylori DNA for cellular proliferation protein #246.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Helicobacter pylori.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 21-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 PA (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR P-PSDB; AAU35933.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX Claim 27; SEQ ID NO 7429; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence encodes an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 927 BP; 314 A; 176 C; 196 G; 241 T; 0 U; 0 Other;
 SQ
 Query Match 74.5%; Score 16.4; DB 4; Length 927;
 Best Local Similarity 94.4%; Pred. No. 4.4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 GGCAGAACTCGGCTAA 21
 DB 249 GGCAGAACTCGGCTAA 232
 RESULT 22
 ABZ38654
 ID ABZ38654 standard; DNA; 1077 BP.
 XX AC ABZ38654;
 XX 07-MAR-2003 (first entry)
 XX N. gonorrhoeae nucleotide sequence SEQ ID 1897.
 DE Antibacterial; infection; vaccine; gene therapy; gene; ds.
 KW Neisseria gonorrhoeae.
 OS WO200279243-A2.
 PN 10-OCT-2002.
 PD 12-FEB-2002; 2002WO-IB002069.
 XX 12-FEB-2001; 2001GB-0000342A.
 PR

XX (CHIR-) CHIRON SPA.
 PA Fontana MR, Piza M, Massignani V, Monaci E;
 PI WPI; 2003-058415/05.
 DR P-PSDB; ABP77684.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection.
 XX Disclosure; Page 325; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
 CC molecules of the invention
 XX Sequence 1077 BP; 235 A; 366 C; 296 G; 180 T; 0 U; 0 Other;
 SQ
 Query Match 74.5%; Score 16.4; DB 7; Length 1077;
 Best Local Similarity 94.4%; Pred. No. 4.5e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CCGGGCAAACTCGGCC 18
 DB 844 CTGGGCAAACTCGGCC 861
 RESULT 23
 ACA48845/C
 ID ACA48845 standard; DNA; 1320 BP.
 XX AC ACA48845;
 XX 19-JUN-2003 (first entry)
 DT Prokaryotic essential gene #30502.
 DE Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX Salmonella paratyphi.
 OS WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR P-PSDB; ABU44975.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 14; SEQ ID NO 36715; 1766pp; English.
 PS

XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1320 BP; 260 A; 364 C; 403 G; 293 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 7; Length 1320;
 Best Local Similarity 94.4%; Pred. No. 4.5e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCCGCC 18
 DB 1237 CCGGCAAAACCTCCGCC 1220

RESULT 24
 ABV76014
 ID ABV76014 standard; cDNA; 3832 BP.

AC ABV76014;

DT 11-FEB-2003 (first entry)

DE Human EGF pathway substrate 8 cDNA.

KW EGF pathway substrate 8; EPS8; human; cancer; diagnosis; cytostatic;
 KW vaccine; gene therapy; adoptive immunotherapy; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 210..2687
 FT /*tag= a
 FT /product= "Human EPS8"

XX WO200208045-A2.

XX 17-OCT-2002.

XX 03-APR-2002; 2002WO-US010439.

XX 04-APR-2001; 2001US-0281566P.

XX (GENZ) GENZYME CORP.

XX Charles AN;
 XX PI
 XX WPI; 2003-075469/07.
 DR P-PSDB; ABP58031.
 DR
 XX Novel composition for inducing immune response in a subject, comprises an
 PT immunogenic ligand individually characterized by its ability to elicit an
 PT immune response against the same native ligand.
 XX
 PS Claim 17; Page 72-73; 76pp; English.

XX The present sequence is that of cDNA encoding human EGF pathway substrate
 CC 8 (EPS8). The invention provides compounds and methods for the detection,
 CC diagnosis and therapy of EPS8-related cancers and malignancies, and for
 CC monitoring the progress of EPS8-related cancers and malignancies.
 CC Immunogenic ligands (see ABP58032-34) and polynucleotides encoding them
 CC (see ABV76015-17) are also provided. A claimed method for inducing an
 CC immune response in a subject involves delivering a composition comprising
 CC at least one of these immunogenic ligands. A claimed method of aiding in
 CC the diagnosis of a neoplastic condition, or susceptibility to a
 CC neoplastic condition, involves determining a level of EPS8 mRNA in a
 CC sample that is at least twice that in a normal or control sample.
 CC detection using a probe or primer based on the present sequence. The
 CC invention also provides compositions which are useful as components of
 CC anti-cancer vaccines and for expanding immune effector cells that are
 CC specific for cancers characterized by expression of EPS8 and which can be
 CC used in adoptive immunotherapy

XX SQ Sequence 3832 BP; 1171 A; 781 C; 840 G; 1040 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 7; Length 3832;
 Best Local Similarity 94.4%; Pred. No. 4.8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCGGCAAAACCTCCGCCCT 19

DB 1164 CCGGCAAAACCTCCACCT 1191

RESULT 25
 ABL03639/c
 ID ABL03639 standard; cDNA; 3963 BP.

XX ABL03639;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5399.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB59536.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.
XX Claim 1; SEQ ID NO 5399; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3963 BP; 924 A; 1087 C; 1143 G; 809 T; 0 U; 0 Other;
SQ
Query Match 74.5%; Score 16.4; DB 4; Length 3963;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCGCTA 20
|||||
DB 3243 GGGCAAAACCTCCGCGCTA 3226
RESULT 26
ID ABL03638 standard; cDNA; 10644 BP.
XX
XX ABL03638;
XX
XX 26-MAR-2002 (first entry)
XX
XX *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 5396.
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX *Drosophila melanogaster*.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB59535.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 5396; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 10644 BP; 3004 A; 2372 C; 2437 G; 2831 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 4; Length 10644;
Best Local Similarity 94.4%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCGCTA 20
|||||
DB 8924 GGGCAAAACCTCCGCGCTA 8907
RESULT 27
ID AAA81476/c
XX AAA81476 standard; DNA; 56485 BP.
XX
XX AAA81476;
XX
XX 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_24 SEQ ID NO:24.
XX
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX *Meningococcus B*; MenB; ds.
XX
XX *Neisseria meningitidis*.
XX
XX WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US023573.
XX
XX 09-OCT-1998; 98US-0103794P.
XX
XX 30-APR-1999; 99US-0132068P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX Rappuoli R, Pizza M;
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
XX in the diagnosis and treatment of N. meningitidis infection and other
XX *Neisseria* infections, for example, N. gonorrhoea.
XX
XX Claim 7; Page 507-524; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic proteins
XX from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent
XX specifically claimed *Neisseria meningitidis* genomic DNA sequences;
XX AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA
XX sequences and their corresponding proteins; AAA81254 to AAA81259 and
XX AAA81304 to AAA81321 represent PCR primers used in the isolation of
XX *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent
XX *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all
XX used in the exemplification of the present invention. The nucleic acid
XX sequences, protein sequences, and antibodies against them, can be used in
XX the manufacture of a composition. The composition can be used as a
XX medicament (or in the manufacture of a medicament) for treating,
XX preventing or diagnosing infection due to *Neisseria* bacteria. For
XX example, some of the identified proteins could be components of vaccines
XX against *Meningococcus B*; against all serotypes; and/or against all
XX pathogenic *Neisseriae*. Identification of sequences from the bacterium
XX will also facilitate production of biological probes, particularly
XX organism-specific probes. Attempts to make efficacious *Meningococcus B*
XX vaccines have failed mainly due to antigen tolerance. Multivalent
XX vaccines have also been tried but none have successfully overcome
XX antigenic variability. The provision of further, complete sequences may
XX provide an opportunity to identify secreted or surface exposed proteins

CC that may be presumed targets for the immune system and which are not
 CC antigenically variable or at least more conserved than other more
 CC variable regions

SQ Sequence 56485 BP; 12504 A; 14247 C; 16158 G; 13573 T; 0 U; 3 Other;

Query Match 74.5%; Score 16.4; DB 3; Length 56485;
 Best Local Similarity 94.4%; Pred. No. 5.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18
 |||||
 DB 35406 CTGGCAAAACCTCGCC 35389

RESULT 28
 AAA81489_4
 Continuation (5 of 9) of AAA81489 from base 400001 (N. meningitidis partial DNA sequence
 WP Sequence split into 9 fragments LOCUS AAA81489 Accession AAA81489

WP	Fragment Name	Begin	End
WP AAA81489_1	1	110000	210000
WP AAA81489_2	100001	210000	310000
WP AAA81489_3	200001	310000	410000
WP AAA81489_4	300001	410000	510000
WP AAA81489_5	400001	510000	610000
WP AAA81489_6	500001	610000	710000
WP AAA81489_7	600001	710000	810000
WP AAA81489_8	700001	810000	837096

Query Match 74.5%; Score 16.4; DB 3; Length 110000;
 Best Local Similarity 94.4%; Pred. No. 5.8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18
 |||||
 DB 71019 CTGGCAAAACCTCGCC 71036

RESULT 29
 AAF21612
 ID AAF21612 standard; DNA; 349980 BP.
 XX
 AC AAF21612;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.

OS Neisseria meningitidis.

XX WO200066791-A1.

PN 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US005928.

XX 30-APR-1999; 99US-0132068P.

PR 08-OCT-1999; 99WO-US023573.

PR 28-FEB-2000; 2000GB-00004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections.

XX Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of Neisseria
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
 CC represent fragments of the NMB genomic sequence, as the sequence was too
 CC long to go in a record on its own it was split into 8 sequences which
 CC overlap each other at the beginning and end of each sequence by 49980 bp
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
 CC primers which are used in the exemplification of the present invention.
 CC The NMB genome and fragments from it have antibacterial activity, and can
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
 CC and/or antibodies which binds to the proteins can be used in compositions
 CC for treating or preventing infection due to Neisserial bacteria or as a
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
 CC computer storage medium or computer databases can be used in a search to
 CC identify open reading frames (ORFs) or coding sequences within the NMB
 CC genome. The DNA sequences provide further opportunities to find antigenic
 CC or immunogenic proteins which are more effective in vaccines than the
 CC outer membrane proteins currently used

XX SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 3; Length 349980;
 Best Local Similarity 94.4%; Pred. No. 6.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18

DB 106248 CTGGCAAAACCTCGCC 106265

RESULT 30

ABZ52802/c

ID ABZ52802 standard; cDNA; 461 BP.

XX AC ABZ52802;

XX 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 1915.

XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;

XX KW expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

XX WO200279476-A1.

XX 10-OCT-2002.

XX 22-MAR-2002; 2002WO-IB000890.

XX 30-MAR-2001; 2001JP-00098371.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (NARE-) NAT RES INST BREWING.

PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

XX Detection of expression of specific Aspergillus genes for monitoring the

PT fermentation and growth conditions of the fungus, using DNA probes.

XX Claim 1; SEQ ID NO 1915; 48pp + Sequence Listing; Japanese.
 PS
 XX
 CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (AB250888-AB256893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridising
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus, rial
 CC especially of Aspergillus oryzae which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 461 BP; 105 A; 134 C; 110 G; 112 T; 0 U; 0 Other;
 Query Match 73.6%; Score 16.2; DB 7; Length 461;
 Best Local Similarity 85.7%; Pred. No. 5.3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCGCCTAAC 22
 Db 96 CGGGCAAAACCTGCGCGACC 76

Search completed: June 20, 2004, 10:17:20
 Job time : 144.538 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1028.06 Seconds

(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22
Sequence: 1 cogggcaaacctccgcctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026579

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
C 1	18.8	85.5	246	12	BJ206364
C 2	18.8	85.5	355	12	BJ176181
C 3	18.8	85.5	365	12	BJ201391
C 4	18.8	85.5	390	12	BJ185747

C 5	18.8	85.5	396	12	BJ185396
C 6	18.8	85.5	406	12	BJ175801
C 7	18.8	85.5	420	12	BJ181492
C 8	18.8	85.5	427	12	BJ183162
C 9	18.8	85.5	503	12	BJ180094
C 10	18.8	85.5	564	12	BJ198230
C 11	18.8	85.5	586	12	BJ183579
C 12	18.8	85.5	666	28	BJ375652
C 13	17.8	80.9	326	10	BE226036
C 14	17.8	80.9	394	12	BE944427
C 15	17.8	80.9	399	10	BF546805
C 16	17.8	80.9	412	10	BE226096
C 17	17.8	80.9	441	9	AI465361
C 18	17.8	80.9	458	9	AI265329
C 19	17.8	80.9	467	10	BE863645
C 20	17.8	80.9	484	9	AA584005
C 21	17.8	80.9	487	14	CD807506
C 22	17.8	80.9	500	14	R48986
C 23	17.8	80.9	510	12	BM950311
C 24	17.8	80.9	532	9	AI598373
C 25	17.8	80.9	606	13	BU059049
C 26	17.8	80.9	658	14	CK134467
C 27	17.8	80.9	697	28	BZ253738
C 28	17.8	80.9	714	13	BQ863377
C 29	17.8	80.9	812	13	BU847656
C 30	17.8	80.9	989	9	AV154279
C 31	17.8	80.9	1101	29	CNS00017
C 32	17.8	80.9	1201	29	CNS016FG
C 33	17.4	79.1	324	28	AQ077340
C 34	17.4	79.1	557	14	CB488692
C 35	17.4	79.1	1036	28	CC268965
C 36	17.4	78.2	427	10	AW596752
C 37	17.2	78.2	431	13	BU577375
C 38	17.2	78.2	513	9	AI736328
C 39	17.2	78.2	513	10	AW836648
C 40	17.2	78.2	533	14	CB352999
C 41	17.2	78.2	537	12	BI974766
C 42	17.2	78.2	559	10	AW655743
C 43	17.2	78.2	628	28	BH065920
C 44	17.2	78.2	782	13	BX880282
C 45	17.2	78.2	790	14	CF45958
C 46	17.2	78.2	916	13	EX414416
C 47	17.2	78.2	959	13	BQ430948
C 48	16.8	76.4	288	12	BI071295
C 49	16.8	76.4	389	12	BG799408
C 50	16.8	76.4	417	13	BY083094
C 51	16.8	76.4	428	10	BF601955
C 52	16.8	76.4	457	9	AV616386
C 53	16.8	76.4	493	10	AW768210
C 54	16.8	76.4	497	29	CG898480
C 55	16.8	76.4	516	12	BI138891
C 56	16.8	76.4	535	13	BU825910
C 57	16.8	76.4	544	14	CB433177
C 58	16.8	76.4	548	9	AV749560
C 59	16.8	76.4	570	14	CF447487
C 60	16.8	76.4	612	29	CG898304
C 61	16.8	76.4	625	29	CG898411
C 62	16.8	76.4	629	14	CG895297
C 63	16.8	76.4	652	29	CG896231
C 64	16.8	76.4	692	29	CG898913
C 65	16.8	76.4	699	29	CG895869
C 66	16.8	76.4	700	29	CG895984
C 67	16.8	76.4	712	29	CG896092
C 68	16.8	76.4	720	13	EX846241
C 69	16.8	76.4	723	29	CG820944
C 70	16.8	76.4	737	29	CG898879
C 71	16.8	76.4	747	14	CF404075
C 72	16.8	76.4	750	28	BZ045990
C 73	16.8	76.4	754	13	BU260771
C 74	16.8	76.4	754	29	CG896553
C 75	16.8	76.4	778	14	CB288874
C 76	16.8	76.4	779	14	CF455372
C 77	16.8	76.4	780	29	CG896489

C 78	16.8	76.4	794	14	CF290814	AGENCYCOURT	C 151	16.4	74.5	1002	10	BE035907	MO16D07 M
C 79	16.8	76.4	847	14	CD301301	AGENCYCOURT	C 152	16.4	74.5	1097	29	CNS04Y0G	AL312361 Tetraodon
C 80	16.8	76.4	869	14	CK156714	FOYDR59TV	153	16.4	74.5	1201	9	AL576960	AL576960 AL576960
C 81	16.8	76.4	876	29	CG822538	CG822538	154	16.4	74.5	1731	29	AV405990	AV405990 Homo sapi
C 82	16.8	76.4	901	13	CG954414	AGENCYCOURT	C 155	16.2	73.6	170	29	CE54586	CE54586 tiger-ges-
C 83	16.8	76.4	916	13	CG954414	AGENCYCOURT	C 156	16.2	73.6	230	9	AA514794	AA514794 ni2d11.s
C 84	16.8	76.4	980	10	BE744910	AGENCYCOURT	C 157	16.2	73.6	237	28	A2899374	A2899374 RPTC-24-2
C 85	16.8	76.4	999	14	CF878999	tr1c018x1	C 158	16.2	73.6	265	9	AV526826	AV526826 AV526826
C 86	16.8	76.4	1180	13	BU959980	AGENCYCOURT	C 159	16.2	73.6	292	10	BF832023	BF832023 PM3-HT090
C 87	16.4	74.5	360	9	AA735226	LD01637.5	C 160	16.2	73.6	303	9	AL838899	AL838899 AL838899
C 88	16.4	74.5	372	28	AZ131094	OSJNB010	C 161	16.2	73.6	314	10	BS550412	BS550412 BS550412
C 89	16.4	74.5	406	14	CF335723	JMT-06-0	C 162	16.2	73.6	351	9	AI188152	AI188152 qd6a02.x
C 90	16.4	74.5	418	28	BE917905	3526_1_58	C 163	16.2	73.6	360	13	CA7675	CA7675 CA7675 Yuj1
C 91	16.4	74.5	456	10	BF481344	FMI_18_H0	C 164	16.2	73.6	364	9	AI626353	AI626353 f04f03.x
C 92	16.4	74.5	486	9	AI938524	sb46G07.Y	C 165	16.2	73.6	367	10	BE118215	BE118215 UI-R-B51
C 93	16.4	74.5	487	10	BB860785	BB860785	C 166	16.2	73.6	383	10	BE711556	BE711556 QV7-HT069
C 94	16.4	74.5	497	28	BH223480	1006113C0	C 167	16.2	73.6	397	14	CD895984	CD895984 G174.1011
C 95	16.4	74.5	505	12	BI315789	saf63d01.	C 168	16.2	73.6	402	28	AF120065	AF120065 AF120065
C 96	16.4	74.5	516	12	BM524868	sa124h12.	C 169	16.2	73.6	419	29	AL943861	AL943861 Arabidops
C 97	16.4	74.5	523	13	EQ173292	EG_CWSL0	C 170	16.2	73.6	421	12	BG669606	BG669606 DRNAH802
C 98	16.4	74.5	529	12	EG688450	336076 BA	C 171	16.2	73.6	425	28	BZ381992	BZ381992 SALK_1176
C 99	16.4	74.5	536	9	AI532051	SD03413.5	C 172	16.2	73.6	428	13	BH0805454	BH0805454 hsa41f04.
C 100	16.4	74.5	537	10	BE474188	sp60809.Y	C 173	16.2	73.6	431	9	AA514432	AA514432 rf61f08.s
C 101	16.4	74.5	540	9	AL694027	DKFZp686E	C 174	16.2	73.6	432	29	AL943863	AL943863 Arabidops
C 102	16.4	74.5	549	13	EX479763	DKFZp686E	C 175	16.2	73.6	446	9	AA916318	AA916318 oh80a01.s
C 103	16.4	74.5	559	10	BE323669	BE323669 sg67410.Y	C 176	16.2	73.6	451	13	BY158980	BY158980 BY158980
C 104	16.4	74.5	561	28	BE917904	3526_1_58	C 177	16.2	73.6	458	28	BH758853	BH758853 SALK_0543
C 105	16.4	74.5	566	9	AL699610	AL699610	C 178	16.2	73.6	460	12	BI708433	BI708433 ft47f05.Y
C 106	16.4	74.5	569	14	CF7091854	DSBFL_24	C 179	16.2	73.6	462	28	BZ241028	BZ241028 BZ241028
C 107	16.4	74.5	571	14	CD035533	mgmt030x0E	C 180	16.2	73.6	465	10	BE129044	BE129044 894021B11
C 108	16.4	74.5	575	14	CD035533	EG_PASL0	C 181	16.2	73.6	466	29	CG665846	CG665846 CG665846
C 109	16.4	74.5	583	12	BF420958	FMI_5.D09	C 182	16.2	73.6	486	28	BZ171343	BZ171343 CH230-369
C 110	16.4	74.5	607	12	EG636165	SD13765.5	C 183	16.2	73.6	489	14	CF3725967	CF3725967 AMGNNUC:S
C 111	16.4	74.5	626	14	CA395805	CG69d11.Y	C 184	16.2	73.6	495	14	R13248	R13248 YF74805.r1
C 112	16.4	74.5	639	28	BZ324059	IC04905.9	C 185	16.2	73.6	496	12	BI360470	BI360470 387419 MA
C 113	16.4	74.5	641	29	CNS04RPN	AL304196 Tetraodon	C 186	16.2	73.6	505	29	CG715396	CG715396 119041E0
C 114	16.4	74.5	641	29	CNS04RPN	AL304196 Tetraodon	C 187	16.2	73.6	506	28	AQ360751	AQ360751 HS_5036.A
C 115	16.4	74.5	675	10	BE473384	sp51e11.Y	C 188	16.2	73.6	511	9	AL840869	AL840869 AL840869
C 116	16.4	74.5	682	28	BH073584	RBC1-24-3	C 189	16.2	73.6	535	28	AZ385344	AZ385344 1M0143H18
C 117	16.4	74.5	691	10	BF496110	AT09824.5	C 190	16.2	73.6	537	14	CF275221	CF275221 NcST3d13
C 118	16.4	74.5	704	10	EG042965	sa446d03.	C 191	16.2	73.6	543	29	TA243G08Q	TA243G08Q T. brucei
C 119	16.4	74.5	709	29	CNS03U10	AL261153 Tetraodon	C 192	16.2	73.6	550	10	B8810043	B8810043 218838 MA
C 120	16.4	74.5	717	9	AI138712	AUI38712	C 193	16.2	73.6	561	14	CF421759	CF421759 NcST3d85
C 121	16.4	74.5	735	29	CG454143	CGVHV72TH	C 194	16.2	73.6	574	29	TA60F11Q	TA60F11Q T. brucei
C 122	16.4	74.5	756	29	CG313413	CGXEW75TH	C 195	16.2	73.6	575	9	AL8383955	AL8383955 AL8383955
C 123	16.4	74.5	756	29	CG313413	CGXEW75TH	C 196	16.2	73.6	577	14	CD3378461	CD3378461 rql2f03.Y
C 124	16.4	74.5	758	12	BI520094	603071221	C 197	16.2	73.6	582	14	CD430889	CD430889 ETH1_5 A0
C 125	16.4	74.5	787	29	CG090617	CGU9C94TH	C 198	16.2	73.6	584	28	AQ651342	AQ651342 Shear3d D
C 126	16.4	74.5	790	29	CG090617	CGU9C94TH	C 199	16.2	73.6	594	9	AV827030	AV827030 AV827030
C 127	16.4	74.5	795	28	BZ542384	OGAKH44TC	C 200	16.2	73.6	601	12	BUI11791	BUI11791 BUI11791
C 128	16.4	74.5	812	14	CD439244	EL01N0522	C 201	16.2	73.6	603	14	CB581925	CB581925 AMGNNUC:N
C 129	16.4	74.5	812	28	BZ814143	PUFEX27TV	C 202	16.2	73.6	605	28	BZ455380	BZ455380 BOMK3CTR
C 130	16.4	74.5	827	29	CG641179	CGM1151TV	C 203	16.2	73.6	609	14	CB580790	CB580790 AMGNNUC:N
C 131	16.4	74.5	864	28	BZ546081	OGAKH44TC	C 204	16.2	73.6	613	14	CB580034	CB580034 AMGNNUC:S
C 132	16.4	74.5	865	28	BZ546081	OGAKH44TC	C 205	16.2	73.6	616	10	BZ658377	BZ658377 BZ658377
C 133	16.4	74.5	890	29	CG457848	PUFLD92TV	C 206	16.2	73.6	621	28	BH685020	BH685020 BOMK3CTR
C 134	16.4	74.5	905	29	CNS013LE	AL103001 Drosophila	C 207	16.2	73.6	631	14	CF261059	CF261059 NcST3c79
C 135	16.4	74.5	907	12	BI460726	603204902	C 208	16.2	73.6	637	12	BM015137	BM015137 603641125
C 136	16.4	74.5	914	28	CG338262	OGI4269TV	C 209	16.2	73.6	638	12	BZ121614	BZ121614 BZ121614
C 137	16.4	74.5	917	28	BZ401922	OGABZ31TC	C 210	16.2	73.6	644	14	CD769015	CD769015 AGENCYCOURT
C 138	16.4	74.5	919	14	CD513277	AGENCYCOURT	C 211	16.2	73.6	648	28	BZ507806	BZ507806 BONGP46TF
C 139	16.4	74.5	921	13	EX451820	EX451820	C 212	16.2	73.6	674	29	CG717809	CG717809 1119050C1
C 140	16.4	74.5	924	13	EX451819	EX451819	C 213	16.2	73.6	703	28	BH936749	BH936749 oq552407.
C 141	16.4	74.5	924	28	CG415868	PUHS843TB	C 214	16.2	73.6	712	13	BX319502	BX319502 BX319502
C 142	16.4	74.5	930	28	BZ814138	PUFEX27TV	C 215	16.2	73.6	712	28	BZ261665	BZ261665 CHX32-326
C 143	16.4	74.5	932	29	CG090614	PUFLD92TV	C 216	16.2	73.6	720	28	CG320840	CG320840 TAM32-22D
C 144	16.4	74.5	948	29	CG361950	OG5B285TC	C 217	16.2	73.6	730	12	BG663088	BG663088 DRA08B01
C 145	16.4	74.5	954	10	BG035260	60232477A	C 218	16.2	73.6	730	28	BH942719	BH942719 oq131912.
C 146	16.4	74.5	985	10	BG035260	60232477A	C 219	16.2	73.6	757	28	CG158837	CG158837 if96d10.g
C 147	16.4	74.5	991	10	BE035753	MO16D08 M	C 220	16.2	73.6	763	29	CG609326	CG609326 OGKAR96TV
C 148	16.4	74.5	996	10	BE035754	MO16D09 M	C 221	16.2	73.6	792	29	CG215679	CG215679 OGGL35TH
C 149	16.4	74.5	997	10	BE035755	MO16D10 M	C 222	16.2	73.6	802	13	BU942525	BU942525 AGENCYCOURT
C 150	16.4	74.5	997	10	BE035755	MO16D11 M	C 223	16.2	73.6	837	28	BH659407	BH659407 BOMBE93TR

297	15.8	71.8	479	12	BJ302954		linear	EST 21-OCT-2003
298	15.8	71.8	483	12	BJ211686			
299	15.8	71.8	484	14	CAY16927	wk4c.pk0		
300	15.8	71.8	486	14	CF332376	NACL--08-		
ALIGNMENTS								
RESULT 1								
BJ206364/c								
LOCUS								
DEFINITION	Bj206364 246 bp mRNA linear EST 21-OCT-2003 caulonemata and rhizoid-like protonemata Physcomitrella patens subsp. patens cDNA clone pphn37d02 5', mRNA sequence.							
ACCESSION	Bj206364							
VERSION	Bj206364.1 GI:18374786							
KEYWORDS	EST.							
SOURCE	Physcomitrella patens subsp. patens							
ORGANISM	Physcomitrella patens subsp. patens							
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariaceae; Physcomitrella. 1 (bases 1 to 246)							
AUTHORS	Nishiyama,T., Fujita,T., Shin-I,T., Seki,M., Nishide,H., Kohara,Y., Kamiya,A., Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohata,Y. and Hasebe,M.							
TITLE	Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution							
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)							
MEDLINE	22709184							
PUBMED	12808149							
COMMENT	Contact: Tadasu Shin-i Center for Genetic Resource Information National Institute of Genetics 1111 Yatae, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a l-PIC phase vector (Carninci et al. 2001) 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector(5'- gagAGGAGGATCCAAACCGTGGAGTTTTTTTTTTTTVN-3', was used as a 1st 3' primer, and 5'-ggTTCGATCATCGTCTTCAGACGCATGACTCGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGCCCAATCGCCGACGTGGAATTGTCGAGACCG'. cDNA instert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003). Protonemata were blended by the POLYTRON, and then cultivated on the BODATG medium for 13- 14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database of Physcomitrella EST clones is available at the PHYSCobase (http://moss.nibb.ac.jp).							
FEATURES	Location/Qualifiers							
source	1..246							
	/organism="Physcomitrella patens subsp. patens"							
	/mol_type="mRNA"							
	/sub_species="patens"							
	/db_xref="taxon:145481"							
	/clone_lib="normalized full length cDNA library,"							
	chloronemata, caulonemata and rhizoid-like protonemata"							
ORIGIN								
Quality Match	85.5%; Score 18.8; DB 12; Length 246;							
Best Local Similarity	90.9%; Pred. No. 4.2e+02;							

Query Match 85.5%; Score 18.8; DB 12; Length 246;
Best Local Similarity 90.9%; Pred. No. 4.2e+02;

Db 172 CCGAGCCAAACCTCCGCTTAC 151

RESULT 4
 BJI185747/c
 LOCUS
 DEFINITION
 BJI185747 390 bp mRNA linear EST 16-OCT-2003
 BJI185747 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb7024 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Physcomitrella patens subsp. patens
 Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella 1 (bases 1 to 390)
 Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution
 Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 22709184
 12808149
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'- GAGAGAGAGAGATCCACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTCGAGTCATCTGTCAGACGATGACTGAGACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGCCAAATCGCCGAGCTCGAATCTCGAGACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database of Physcomitrella EST clones is available at the PHYSCoBase (http://moss.nibb.ac.jp).

FEATURES
 source
 1..390
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb7024"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"

ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 390;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGCGAAACCTCCGCTTAC 22
 |||
 162 CCGAGCCAAACCTCCGCTTAC 141

Db

RESULT 5
 BJI185396/c
 LOCUS
 DEFINITION
 BJI185396 396 bp mRNA linear EST 16-OCT-2003
 BJI185396 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb6021 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Physcomitrella patens subsp. patens
 Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella 1 (bases 1 to 396)
 Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution
 Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 22709184
 12808149
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'- GAGAGAGAGAGATCCACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTCGAGTCATCTGTCAGACGATGACTGAGACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGCCAAATCGCCGAGCTCGAATCTCGAGACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database of Physcomitrella EST clones is available at the PHYSCoBase (http://moss.nibb.ac.jp).

FEATURES
 source
 1..396
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb6021"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"

ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 396;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGCGAAACCTCCGCTTAC 22
 |||
 204 CCGAGCCAAACCTCCGCTTAC 183

Db

RESULT 6

CDNA clone pphb36107 5', mRNA sequence.

CESSION
 BU183162
 VERSION
 BU183162.1 GI:18351111
 EST.

Phycomitrella patens subsp. patens
 Phycomitrella patens subsp. patens
 Bryophyta; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 1 (bases 1 to 427)

REFERENCE
 Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H.,
 Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K.,
 Kohara, Y. and Hasebe, M.
 Comparative genomics of Physcomitrella patens gametophytic
 transcriptome and Arabidopsis thaliana: Implication for land plant
 evolution

TITLE
 Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Contact: Tadasu Shin-i
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript II (KS),
 that was in vivo excised from a 1-FLC phase vector (Carninci et al.
 2001). 5' end of the cDNA that was digested with XhoI was ligated
 to SalI site of the vector and the 3' end including polyA tail was
 ligated to BamHI site of the
 vector (5'- GAGAGAGAGATCAACCCCTGGAGAGTCTTTTTTTTTTTT-3' was
 used as a 1st 3' primer, and
 5'-GGTTCGAGTCATCGCTGTCAGACACGATGACTCGAGACCGNNNN-3' as 2nd
 5'-hairpin primer, giving the following 5' boarder sequence,
 AGCCAAATCGCCGAGCTCGAATTCGTGGAACCG). cDNA insert could be
 amplified with conventional T7 and T3 primers. This full-length
 cDNA library was generated according to the method described in
 Nishiyama et al. (2003).

Protonemata were blended by the POLYTRON, and then cultivated on
 the BCDATG medium for 13- 14 days under the continuous light.
 These clones are available from RIKEN Bio Resource Center.
 (http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
 of Physcomitrella EST clones is available at the PHYSCOBASE
 (http://moss.nibb.ac.jp).

FEATURES
 source
 1..427
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clones="pphb36107"
 /tissue_type="mixture of chloronemata, caulonemata and
 malformed buds"
 /clone_lib="normalized full length cDNA library,
 chloronemata, caulonemata and malformed buds"

IGIN

Query Match 85.5%; Score 18.8; DB 12; Length 427;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 CCGGCGAAACCTCGGCTAAC 22
 |||
 166 CCGAGCCAAACCTCGGCTAAC 145

RESULT 9
 180094/c
 CCUS
 DEFINITION
 caulonemata and malformed buds Physcomitrella patens subsp. patens
 cDNA clone pphb27c19 5', mRNA sequence.
 BU180094

REFERENCE 1 (bases 1 to 666)
 AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
 MEDLINE 22542063
 PUBMED 12655398
 COMMENT Other GSSs: AG-ND-178P24.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.
 Location/Qualifiers
 1. 666
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-178P24"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

FEATURES
 source
 1. 666

ORIGIN
 Query Match 85.5%; Score 18.8; DB 28; Length 666;
 Best Local Similarity 90.9%; Pred. No. 4.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCCTAAC 22
 |||||
 Db 521 CCGGGCAAAACCTCCGCCTAAC 542

RESULT 13
 BE226036
 LOCUS
 DEFINITION iai3a09.v1 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA clone IMAGE:5638168 5', similar to SM.TPMZ_RAT P18344 TROPOMYOSIN ALPHA CHAIN, BRAIN-3 ;, mRNA sequence.

ACCESSION BE226036
 VERSION BE226036.1 GI:8931272
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 326)
 Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
 WashU-Harvard Pancreas EST Project
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

REFERENCE 1 (bases 1 to 666)
 AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
 MEDLINE 22542063
 PUBMED 12655398
 COMMENT Other GSSs: AG-ND-178P24.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.
 Location/Qualifiers
 1. 666
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-178P24"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

FEATURES
 source
 1. 666

ORIGIN
 Query Match 85.5%; Score 18.8; DB 28; Length 666;
 Best Local Similarity 90.9%; Pred. No. 4.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCCTAAC 22
 |||||
 Db 521 CCGGGCAAAACCTCCGCCTAAC 542

RESULT 13
 BE226036
 LOCUS
 DEFINITION iai3a09.v1 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA clone IMAGE:5638168 5', similar to SM.TPMZ_RAT P18344 TROPOMYOSIN ALPHA CHAIN, BRAIN-3 ;, mRNA sequence.

ACCESSION BE226036
 VERSION BE226036.1 GI:8931272
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 326)
 Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
 WashU-Harvard Pancreas EST Project
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Adimika Meadows (meadows@fas.harvard.edu)
 MGI:1847192 this sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 323.

FEATURES

source
 1. 326
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone="IMAGE:5638168"
 /sex="Both"
 /tissue_type="Pancreatic Bud"
 /dev_stage="Embryonic day 10.5 and 12.5, mixed"
 /lab_host="DH10B"
 /clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
 /note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.47 kb. Primary library, unamplified. cDNA Library Preparation: Guolin Chen."

ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 326;
 Best Local Similarity 90.5%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGGCAAAACCTCCGCCTAAC 22
 |||||
 Db 73 CGGGCAAAACCTCCGCCTAAC 93

RESULT 14

BE944427

LOCUS

DEFINITION

UI-M-EHOp-bvr-c-22-0-UI.r1 NIH_BMAP_EHOp Mus musculus cDNA clone IMAGE:5695821 5', mRNA sequence.

ACCESSION

BE944427

VERSION

BE944427.1 GI:19428012

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 394)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

source

1. 394
 /organism="Mus musculus"

Seq primer: PYX-5.

Location/Qualifiers

1. 394

/organism="Mus musculus"

```

/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5695821"
/tissue_type="whole brain"
/lab_host="embryo 18.5 dpc"
/clone_lib="NIH BMAP EH0p"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

```

Query Match 80.9%; Score 17.8; DB 12; Length 394;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTTAAC 22
DB 70 CGAGCAAAACCTCCGCTTAAC 90

```

RESULT 15
BF546805
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

```

BF546805 399 bp mRNA linear EST 11-DEC-2000
UI-R-C2p-qv-f-01-0-UI-r1 UI-R-C2p Rattus norvegicus cDNA clone
UI-R-C2p-qv-f-01-0-UI-5', mRNA sequence.
BF546805
EST.
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 399)
Bernaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
9704447
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LNLN (info@image.llnl.gov). IMAGE ID= 1789368 The following repetitive elements were found in this cDNA sequence: 54-145, >BI-FHSINE/Alu
Seq primer: M13 Forward.
Location/Qualifiers
1. .399
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"

```

/db_xref="taxon:10116"
/clone="UI-R-C2p-qv-f-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C2p"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bernaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

```

Query Match 80.9%; Score 17.8; DB 10; Length 399;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTTAAC 22
DB 322 CGGCAAAACCTCCGCTTAAC 342

```

RESULT 16
BE226096
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

BE226096 412 bp mRNA linear EST 13-MAR-2002
1a14a08.Y1 Mouse E10 5 12 5 Pancreas cDNA library Mus musculus cDNA clone IMAGE:5638262 5', similar to SW.TPMZ_RAT P18344 TROPOMYOSIN ALPHA CHAIN, BRAIN-3 ;, mRNA sequence.
BE226096
BE226096.1 GI:8931332
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 412)
Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, V.
WashU-Harvard Pancreas EST Project
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohph.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Adimika Meadows (meadows@fas.harvard.edu)
MGI:1847286 This sequence now available from the IMAGE consortium,

for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 403.
 Location/Qualifiers
 1..412
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone="IMAGE:5638262"
 /sex="Both"
 /tissue_type="Pancreatic Bud"
 /dev_stage="Embryonic day 10.5 and 12.5, mixed"
 /lab_host="DHI08"
 /clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
 /note="Organ: Pancreas; Vector: pSPOR1; Site 1: Not 1;
 Site 2: Sal 1; Library constructed using SuperScript
 Plasmid library kit (Life Technologies). cDNA made by
 oligo-dT priming. Size-selected by column fractionation;
 average insert size 1.47 kb. Primary library, unamplified.
 cDNA Library Preparation: Guolin Chen."

FEATURES

source

ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 412;
 Best Local Similarity 90.5%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCGAAACCTCGCCTAAC 22
 |||||
 Db 73 CGAGCAAAACCGCGCCTAAC 93

RESULT 17

AI465361/c 441 bp mRNA linear EST 09-MAR-1999
 LOCUS mu75903.x1 Stratagene mouse embryonic carcinoma (#937317) Mus
 DEFINITION tropomyosin alpha chain, BRAIN-3, mRNA sequence.
 ACCESSION AI465361 GI:4319391
 VERSION 1
 KEYWORDS Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 441)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Theising, B., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 396.

FEATURES

source

/clone_lib="Stratagene mouse embryonic carcinoma
 (#937317)"
 /note="vector: pbluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; Cloned unidirectionally. Primer: Oligo dr. P19 cell
 line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
 adaptor sequence: 5' GAATTCGGACGAG 3' ~3' adaptor
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 441;
 Best Local Similarity 90.5%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCGAAACCTCGCCTAAC 22
 |||||
 Db 352 CGAGCAAAACCGCGCCTAAC 332

RESULT 18

AI265329/c 458 bp mRNA linear EST 18-NOV-1998
 LOCUS uk01908.y1 Schiller mouse MAC13 Mus musculus cDNA clone
 DEFINITION IMAGE:1958174 5' similar to gp:M19267 TROPOMYOSIN, FIBROBLAST
 ISOFORM TM3 (HUMAN); gb:X64831 M.musculus mRNA for skeletal muscle
 alpha tropomyosin (MOUSE);, mRNA sequence.
 ACCESSION AI265329
 VERSION 1
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus

REFERENCE

1 (bases 1 to 458)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Theising, B., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1984514
 Seq primer: -40RP from Gibco
 High quality sequence stop: 431.

FEATURES

source

1..458
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:1958174"
 /tissue_type="colon cancer"
 /cell_line="colon cancer cell line MAC13"
 /lab_host="SOLR"
 /clone_lib="Schiller mouse MAC13"
 /note="Vector: pBluescript SK- (Stratagene); Site 1:
 EcoRI; Site 2: XhoI; Double-stranded cDNA was prepared
 from cell line MAC13 using primer
 5'-GAGAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
 adaptor was used on the 5' end of the cDNA as follows:
 5'-AATTCGGACGAG-3'. The library was size-selected and
 went through one round of amplification. Average insert
 size is 1.7 kb, with a range from 0.4-12 kb. This library
 was constructed by Dr. Martin Schiller (Johns Hopkins
 University)."

ORIGIN

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: csapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source

Seq primer: pyX-5.
Location/Qualifiers
1. .487
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:3054291"
/tissue_type="whole eye"
/dev_stage="embryo 15.5,16.5,17.5,18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GW0"
/note="Organ: Eye; Vector: pyX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pyX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CTGGTCTC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match 80.9%; Score 17.8; DB 14; Length 487;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGCGAAACCTCGCCTTAC 22

Db 167 CGAGCAAAACCGCGCCTTAC 187

RESULT 22

R48986/c
LOCUS R48986 500 bp mRNA linear EST 22-MAY-1995
DEFINITION Y95501.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:38159 3', mRNA sequence.

ACCESSION R48986

VERSION R48986.1 GI:817750

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 500)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

CONTACT: Wilson RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1350

High quality sequence stops: 326 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1350 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 326.

FEATURES
source

1. .500
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:410700"
/db_xref="taxon:9606"
/clone="IMAGE:38159"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="CH10B (ampicillin resistant)"
/clone_lib="Soares infant brain INIB"
/note="Organ: whole brain; Vector: Lfamid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo (dT) primer [5',
RACGGAAGAAATCGCGCCGAGCAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 14; Length 500;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGCCTTAA 21

Db 395 CCGGGCAAAACCTCGCCTTGA 375

RESULT 23

BM950311

LOCUS BM950311

DEFINITION

UI-X-EHOP-but-p-21-0-UI-r1 NIH_BMAP_EHOP Mus musculus cDNA clone

IMAGE:5686916 5', mRNA sequence.

ACCESSION BM950311

VERSION BM950311.1 GI:19433901

KEYWORDS EST.

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 510)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: csapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pyX-5.

Location/Qualifiers

1. .510

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

ACCESSION CK134467
VERSION CK134467.1 GI:38626403
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 658)
AUTHORS Scapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Friese, E., George, R., Gonzalez, M., Guatin, E., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Other ESTs: RB4552.Sprime
Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003486: arm:X [11180449,11484096]
estimated-cyto:10B11-10D4: 01/30/2002
Plate: RE.645 row: E column: 4
High quality sequence stop: 480.
Location/Qualifiers
1..658
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RB64552"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DMS-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFic-1"
/note="Organ: embryo; Vector: pFic1; Site:1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
Query Match 80.9%; Score 17.8; DB 14; Length 658;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGGCAAAACCTCCGCTAA 21
Db 504 CCGGGCATAACCTCCGCTGA 524
RESULT 27
BZ253738
LOCUS BZ253738 697 bp DNA linear GSS 12-OCT-2002
DEFINITION CH230-306P22.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-306P22, genomic survey sequence.
ACCESSION BZ253738
VERSION BZ253738.1 GI:23915075
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Teegave, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment

JOURNAL COMMENT
Unpublished (1999)
Other GSSs: CH230-306P22.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm). BAC end page: http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html
Plate: 306 row: p column: 22
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..697
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strains="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-306P22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN
Query Match 80.9%; Score 17.8; DB 28; Length 697;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGGCAAAACCTCCGCTAA 21
Db 468 CCGGGCAAAACATCCCTTAA 488
RESULT 28
BQ863377/c
LOCUS BQ863377 714 bp mRNA linear EST 14-AUG-2002
DEFINITION QGC23LI2.Y9.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
QGC23LI2, mRNA sequence.
ACCESSION BQ863377
VERSION BQ863377.1 GI:22248842
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 714)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compgenomics.ucdavis.edu/>
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmudson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA.Contig2235, see <http://cgfdb.ucdavis.edu/> for details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

source
 1. .989
 location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="2900060L06"
 /sex="male"
 /tissue_type="hippocampus"
 /dev_stage="adult"
 /clone_lib="Mus musculus hippocampus C57BL/6J adult"

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 989;
 Best Local Similarity 90.5%; Pred. No. 1.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCGGCTAAC 22
 || ||||| |||||
 Db 296 CGAGCAAAACCGCGGCTAAC 316

Search completed: June 20, 2004, 14:14:19
 Job time : 1055.06 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 24.328 Seconds

(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 ccggcgaacacccgcctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA:**

1: /cgn2_6/prodata/2/ina/5A COMB.seq:**

2: /cgn2_6/prodata/2/ina/5B COMB.seq:**

3: /cgn2_6/prodata/2/ina/6A COMB.seq:**

4: /cgn2_6/prodata/2/ina/6B COMB.seq:**

5: /cgn2_6/prodata/2/ina/PCUS COMB.seq:**

6: /cgn2_6/prodata/2/ina/backfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	4403765	3	US-09-103-840A-2
C 2	22	100.0	4411529	3	US-09-103-840A-1
C 3	15.8	71.8	954	3	US-08-680-506-5
C 4	15.8	71.8	1347	3	US-08-680-506-8
C 5	15.8	71.8	1587	3	US-08-680-506-6
C 6	15.8	71.8	2233	3	US-08-680-506-4
C 7	15.8	71.8	5027	3	US-08-680-506-2
C 8	15.6	70.9	876	4	US-09-328-352-1452
C 9	15.6	70.9	1041	4	US-09-266-965-38
C 10	15.6	70.9	1206	4	US-09-489-039A-3575
C 11	15.6	70.9	2044	6	5229279-3
C 12	15.6	70.9	2094	6	5512669-3
C 13	15.6	70.9	19513	4	US-10-204-708-40
C 14	15.6	70.9	31940	4	US-09-596-002-13
C 15	15.6	70.9	53500	4	US-09-366-965-76
C 16	15.2	69.1	273	4	US-09-489-039A-2975
C 17	15.2	69.1	447	4	US-09-489-039A-2954
C 18	15.2	69.1	534	4	US-09-107-532A-2574
C 19	15.2	69.1	1588	4	US-09-976-594-625
C 20	15.2	69.1	1791	4	US-09-489-039A-2299
C 21	15.2	69.1	1830121	4	US-09-557-884-1
C 22	15.2	69.1	1830121	4	US-09-643-990A-1
C 23	15	68.2	6642	1	US-08-727-034-5
C 24	15	68.2	6843	1	US-08-727-034-6
C 25	14.8	67.3	774	4	US-08-858-207A-177
C 26	14.8	67.3	1038	4	US-09-328-352-715
C 27	14.8	67.3	1206	4	US-09-489-039A-1114

C 28	14.8	67.3	1482	4	US-09-543-681A-445	Sequence 445, App
C 29	14.8	67.3	1660	4	US-09-620-312D-723	Sequence 723, App
C 30	14.8	67.3	1770	4	US-09-489-039A-1224	Sequence 1224, App
C 31	14.8	67.3	4597	4	US-09-221-017B-1093	Sequence 1093, App
C 32	14.8	67.3	5004	4	US-09-489-039A-4453	Sequence 4453, App
C 33	14.8	67.3	5467	4	US-09-976-594-735	Sequence 735, App
C 34	14.8	67.3	28171	4	US-08-961-527-22	Sequence 22, Appli
C 35	14.8	67.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 36	14.8	67.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 37	14.6	66.4	63	3	US-07-921-104D-32	Sequence 32, Appli
C 38	14.6	66.4	345	4	US-09-489-039A-1576	Sequence 1576, Ap
C 39	14.6	66.4	351	3	US-07-921-104D-71	Sequence 71, Appl
C 40	14.6	66.4	375	3	US-07-921-104D-30	Sequence 30, Appl
C 41	14.6	66.4	506	4	US-09-621-976-3772	Sequence 2772, Ap
C 42	14.6	66.4	1203	4	US-09-489-039A-3904	Sequence 3904, Ap
C 43	14.6	66.4	1254	4	US-09-489-039A-4840	Sequence 4840, Ap
C 44	14.6	66.4	1488	4	US-09-252-931A-9307	Sequence 9307, Ap
C 45	14.6	66.4	1538	4	US-09-976-594-52	Sequence 52, Appl
C 46	14.6	66.4	1557	4	US-09-489-039A-4776	Sequence 4776, Ap
C 47	14.6	66.4	1569	4	US-09-724-623-57	Sequence 57, Appl
C 48	14.6	66.4	1602	4	US-09-252-931A-9270	Sequence 9270, Ap
C 49	14.6	66.4	1990	4	US-09-976-594-53	Sequence 53, Appl
C 50	14.6	66.4	2061	4	US-09-543-681A-1304	Sequence 1304, Ap
C 51	14.6	66.4	2992	4	US-08-221-017B-819	Sequence 819, App
C 52	14.6	66.4	3207	1	US-08-162-081B-35	Sequence 35, Appl
C 53	14.6	66.4	3207	2	US-08-780-872-35	Sequence 35, Appl
C 54	14.6	66.4	3207	3	US-09-085-957-35	Sequence 35, Appl
C 55	14.6	66.4	3240	1	US-08-162-081B-34	Sequence 34, Appl
C 56	14.6	66.4	3240	2	US-08-780-872-34	Sequence 34, Appl
C 57	14.6	66.4	3240	3	US-08-085-957-34	Sequence 34, Appl
C 58	14.6	66.4	3412	1	US-08-162-081B-32	Sequence 32, Appl
C 59	14.6	66.4	3412	2	US-08-780-872-32	Sequence 32, Appl
C 60	14.6	66.4	3412	3	US-09-085-957-32	Sequence 32, Appl
C 61	14.6	66.4	9432	1	US-08-277-231A-1	Sequence 1, Appli
C 62	14.6	66.4	9432	2	US-08-473-750-4	Sequence 4, Appli
C 63	14.6	66.4	9432	2	US-08-477-326-4	Sequence 4, Appli
C 64	14.4	65.5	250	4	US-08-956-171E-4568	Sequence 4568, Ap
C 65	14.4	65.5	252	4	US-09-134-001C-2253	Sequence 2253, Ap
C 66	14.4	65.5	356	4	US-08-956-171E-4293	Sequence 4293, Ap
C 67	14.4	65.5	502	4	US-08-956-171E-3907	Sequence 3907, Ap
C 68	14.4	65.5	609	4	US-08-956-171E-3937	Sequence 3937, Ap
C 69	14.4	65.5	649	4	US-09-470-191-5	Sequence 5, Appli
C 70	14.4	65.5	770	3	US-08-998-416-63	Sequence 63, Appl
C 71	14.4	65.5	969	4	US-09-134-001C-2283	Sequence 2283, Ap
C 72	14.4	65.5	1290	4	US-08-956-171E-3194	Sequence 3194, Ap
C 73	14.4	65.5	2047	3	US-08-836-261A-1	Sequence 1, Appli
C 74	14.4	65.5	4121	4	US-09-638-524A-1	Sequence 1, Appli
C 75	14.4	65.5	5476	4	US-10-204-708-81	Sequence 81, Appl
C 76	14.4	65.5	5666	4	US-10-204-708-30	Sequence 30, Appl
C 77	14.4	65.5	6156	4	US-10-204-708-60	Sequence 60, Appl
C 78	14.4	65.5	6343	3	US-08-581-148C-30	Sequence 30, Appl
C 79	14.2	64.5	384	4	US-09-489-039A-6139	Sequence 6139, Ap
C 80	14.2	64.5	462	4	US-09-252-931A-13639	Sequence 13639, A
C 81	14.2	64.5	471	4	US-09-489-039A-289	Sequence 289, App
C 82	14.2	64.5	501	4	US-09-252-931A-9251	Sequence 9251, Ap
C 83	14.2	64.5	513	4	US-09-489-039A-2764	Sequence 2764, Ap
C 84	14.2	64.5	516	4	US-09-252-931A-6461	Sequence 6461, Ap
C 85	14.2	64.5	525	4	US-09-252-931A-8915	Sequence 8915, Ap
C 86	14.2	64.5	525	4	US-09-489-039A-335	Sequence 335, App
C 87	14.2	64.5	552	4	US-09-489-039A-4449	Sequence 4449, Ap
C 88	14.2	64.5	594	4	US-09-252-931A-13801	Sequence 13801, A
C 89	14.2	64.5	603	4	US-09-252-931A-2672	Sequence 2672, Ap
C 90	14.2	64.5	621	4	US-09-252-931A-6380	Sequence 6380, Ap
C 91	14.2	64.5	714	4	US-09-252-931A-8702	Sequence 8702, Ap
C 92	14.2	64.5	735	4	US-08-561-527-366	Sequence 366, App
C 93	14.2	64.5	786	4	US-09-252-931A-2865	Sequence 2865, Ap
C 94	14.2	64.5	823	3	US-08-998-416-493	Sequence 493, App
C 95	14.2	64.5	840	4	US-09-489-039A-6039	Sequence 6039, Ap
C 96	14.2	64.5	873	4	US-09-252-931A-7015	Sequence 7015, Ap
C 97	14.2	64.5	906	4	US-09-252-931A-8799	Sequence 8799, Ap
C 98	14.2	64.5	933	4	US-09-252-931A-6169	Sequence 6169, Ap
C 99	14.2	64.5	1068	4	US-09-711-164-246	Sequence 246, App
C 100	14.2	64.5	1104	4	US-09-134-001C-2720	Sequence 2720, Ap

101	14.2	64.5	1203	4	US-09-252-991A-8410	Sequence 8410, App	174	14	63.6	1746	4	US-09-252-991A-13461	Sequence 13461, A
102	14.2	64.5	1221	4	US-09-252-991A-8296	Sequence 8296, App	175	14	63.6	1830	4	US-09-252-991A-8322	Sequence 8322, App
103	14.2	64.5	1239	4	US-09-489-039A-5782	Sequence 5782, App	176	14	63.6	2207	3	US-08-956-322-3	Sequence 3, Appli
C 104	14.2	64.5	1347	4	US-09-107-532A-308	Sequence 308, App	177	14	63.6	2210	3	US-08-464-700-53	Sequence 53, Appli
C 105	14.2	64.5	1350	4	US-09-252-991A-6012	Sequence 6012, App	178	14	63.6	2429	2	US-08-956-322-1	Sequence 1, Appli
C 106	14.2	64.5	1371	4	US-09-489-039A-1161	Sequence 1161, App	179	14	63.6	2427	2	US-08-678-039A-39	Sequence 39, Appli
C 107	14.2	64.5	1494	4	US-09-252-991A-13441	Sequence 13441, A	180	14	63.6	2436	4	US-09-199-637A-92	Sequence 92, Appli
C 108	14.2	64.5	1523	4	US-09-673-995A-1342	Sequence 1342, App	181	14	63.6	2535	4	US-09-672-737A-10	Sequence 10, Appli
C 109	14.2	64.5	1533	4	US-09-252-991A-800	Sequence 800, App	182	14	63.6	2730	1	US-08-344-536-1	Sequence 1, Appli
C 110	14.2	64.5	1569	4	US-09-252-991A-3005	Sequence 3005, App	183	14	63.6	2730	3	US-08-920-582-1	Sequence 1, Appli
C 111	14.2	64.5	1680	2	US-08-759-581B-3	Sequence 3, Appli	184	14	63.6	2748	4	US-09-252-991A-13820	Sequence 13820, A
C 112	14.2	64.5	1680	4	US-09-304-711-3	Sequence 3, Appli	185	14	63.6	3350	3	US-09-110-116-2	Sequence 2, Appli
C 113	14.2	64.5	1680	4	US-09-173-281-3	Sequence 3, Appli	186	14	63.6	3720	4	US-09-252-991A-13627	Sequence 13627, A
C 114	14.2	64.5	1701	4	US-09-252-991A-1960	Sequence 1960, App	187	14	63.6	4421	2	US-08-257-963B-9	Sequence 9, Appli
C 115	14.2	64.5	1752	4	US-09-543-681A-1651	Sequence 1651, App	188	14	63.6	4421	2	US-08-367-841A-9	Sequence 9, Appli
C 116	14.2	64.5	1761	4	US-09-489-039A-1779	Sequence 1779, App	189	14	63.6	4421	4	US-08-520-373D-6	Sequence 6, Appli
C 117	14.2	64.5	1770	4	US-09-107-532A-17432	Sequence 17432, App	190	14	63.6	4421	5	PCT-US95-07201-9	Sequence 9, Appli
C 118	14.2	64.5	1788	4	US-09-252-991A-863	Sequence 863, App	191	14	63.6	6909	4	US-09-199-637A-111	Sequence 111, App
C 119	14.2	64.5	1804	1	US-08-306-691B-40	Sequence 40, Appli	192	14	63.6	8931	3	US-09-028-934-28	Sequence 28, Appli
C 120	14.2	64.5	1804	4	US-09-167-322-14	Sequence 14, Appli	193	14	63.6	9862	4	US-09-691-861A-3	Sequence 3, Appli
C 121	14.2	64.5	1804	5	PCT-US93-06251-82	Sequence 82, Appli	194	14	63.6	24707	4	US-09-740-037-3	Sequence 3, Appli
C 122	14.2	64.5	1845	4	US-09-489-039A-2790	Sequence 2790, App	195	14	63.6	42235	4	US-09-199-637A-1	Sequence 1, Appli
C 123	14.2	64.5	1935	4	US-09-252-991A-8438	Sequence 8438, App	196	14	63.6	45819	4	US-08-453-702B-72	Sequence 72, Appli
C 124	14.2	64.5	1977	4	US-09-252-991A-8339	Sequence 8339, App	197	14	63.6	11282	4	US-09-754-250-3	Sequence 3, Appli
C 125	14.2	64.5	2097	4	US-09-252-991A-885	Sequence 885, App	198	13.8	62.7	38	4	US-09-371-772B-12522	Sequence 12522, A
C 126	14.2	64.5	2098	4	US-09-489-847-20	Sequence 20, Appli	199	13.8	62.7	48	4	US-09-453-702B-153	Sequence 153, App
C 127	14.2	64.5	2169	4	US-09-489-847-20	Sequence 1910, App	200	13.8	62.7	254	4	US-09-016-434-743	Sequence 743, App
C 128	14.2	64.5	2230	3	US-09-252-991A-1910	Sequence 7, Appli	201	13.8	62.7	295	4	US-09-313-294A-4988	Sequence 4988, App
C 129	14.2	64.5	2230	4	US-08-448-194-7	Sequence 1, Appli	202	13.8	62.7	321	4	US-09-313-294A-6928	Sequence 6928, App
C 130	14.2	64.5	2260	2	US-08-788-750-1	Sequence 1, Appli	203	13.8	62.7	402	4	US-09-634-238-96	Sequence 96, Appli
C 131	14.2	64.5	2298	4	US-09-252-991A-13540	Sequence 13540, A	204	13.8	62.7	484	4	US-09-328-475C-183	Sequence 183, App
C 132	14.2	64.5	2397	4	US-09-489-039A-2874	Sequence 2874, App	205	13.8	62.7	526	3	US-09-328-111-361	Sequence 261, App
C 133	14.2	64.5	2412	4	US-09-252-991A-7043	Sequence 7043, App	206	13.8	62.7	635	3	US-09-328-111-56	Sequence 56, Appli
C 134	14.2	64.5	2412	4	US-09-252-991A-13908	Sequence 13908, A	207	13.8	62.7	714	3	US-08-971-090-2	Sequence 2, Appli
C 135	14.2	64.5	2472	4	US-09-252-991A-7084	Sequence 7084, App	208	13.8	62.7	913	3	US-08-971-090-3	Sequence 3, Appli
C 136	14.2	64.5	2916	4	US-09-976-594-210	Sequence 210, App	209	13.8	62.7	734	4	US-09-252-991A-8606	Sequence 8606, App
C 137	14.2	64.5	2946	4	US-09-252-991A-2045	Sequence 2045, App	210	13.8	62.7	1089	4	US-09-252-991A-1162	Sequence 1162, App
C 138	14.2	64.5	3286	2	US-08-363-124A-1	Sequence 1, Appli	211	13.8	62.7	1125	4	US-08-107-532A-2373	Sequence 2373, App
C 139	14.2	64.5	3509	2	US-08-817-436A-1	Sequence 1, Appli	212	13.8	62.7	1131	6	5180810-3	Patent No. 5180810
C 140	14.2	64.5	5228	4	US-10-164-595-23	Sequence 23, Appli	213	13.8	62.7	1146	4	US-09-252-991A-8577	Sequence 8577, App
C 141	14.2	64.5	9278	1	US-08-243-542-9	Sequence 9, Appli	214	13.8	62.7	1182	4	US-09-543-681A-817	Sequence 817, App
C 142	14.2	64.5	9278	1	US-08-477-407-9	Sequence 9, Appli	215	13.8	62.7	1199	4	US-08-630-915A-15	Sequence 15, Appli
C 143	14.2	64.5	9278	1	US-08-484-355-9	Sequence 9, Appli	216	13.8	62.7	1260	4	US-09-328-352-848	Sequence 848, App
C 144	14.2	64.5	10357	4	US-08-961-527-191	Sequence 191, App	217	13.8	62.7	1326	4	US-09-328-352-3628	Sequence 3628, App
C 145	14.2	64.5	18912	5	PCT-US96-03916-59	Sequence 191, App	218	13.8	62.7	1364	4	US-09-620-322D-36	Sequence 36, Appli
C 146	14.2	64.5	36063	4	US-08-311-731A-140	Sequence 59, Appli	219	13.8	62.7	1377	4	US-09-252-991A-11683	Sequence 11683, A
C 147	14.2	64.5	42157	4	US-08-311-731A-126	Sequence 126, App	220	13.8	62.7	1400	1	US-08-051-142-1	Sequence 1, Appli
C 148	14.2	64.5	42157	4	US-08-311-731A-13427	Sequence 126, App	221	13.8	62.7	1434	4	US-09-252-991A-207	Sequence 207, App
C 149	14.2	64.5	42157	4	US-08-356-575E-22	Sequence 22, Appli	222	13.8	62.7	1446	4	US-09-489-039A-5151	Sequence 5151, App
C 150	14.2	64.5	315	3	US-09-252-991A-13101	Sequence 13101, A	223	13.8	62.7	1506	4	US-09-252-991A-224	Sequence 224, App
C 151	14.2	64.5	387	4	US-09-073-009-133	Sequence 133, App	224	13.8	62.7	1536	4	US-09-252-991A-11555	Sequence 11555, A
C 152	14.2	64.5	387	4	US-09-073-010-133	Sequence 133, App	225	13.8	62.7	1560	1	US-07-813-584A-1	Sequence 1, Appli
C 153	14.2	64.5	405	4	US-09-252-991A-13883	Sequence 13883, A	226	13.8	62.7	1560	1	US-08-330-515-1	Sequence 1, Appli
C 154	14.2	64.5	414	4	US-09-252-991A-13427	Sequence 13427, A	227	13.8	62.7	1613	3	US-08-725-594A-32	Sequence 32, Appli
C 155	14.2	64.5	546	4	US-09-199-637A-98	Sequence 98, Appli	228	13.8	62.7	1613	3	US-08-937-993-32	Sequence 32, Appli
C 156	14.2	64.5	608	4	US-09-221-017B-456	Sequence 456, App	229	13.8	62.7	1674	4	US-09-252-991A-1253	Sequence 1253, App
C 157	14.2	64.5	615	4	US-09-489-039A-375	Sequence 375, App	230	13.8	62.7	1784	6	5180810-2	Patent No. 5180810
C 158	14.2	64.5	713	3	US-09-252-991A-13717	Sequence 53, Appli	231	13.8	62.7	1923	4	US-09-503-814A-5	Sequence 5, Appli
C 159	14.2	64.5	798	4	US-09-252-991A-13717	Sequence 13717, A	232	13.8	62.7	1923	4	US-09-503-814A-7	Sequence 7, Appli
C 160	14.2	64.5	809	3	US-09-252-991A-13717	Sequence 27, Appli	233	13.8	62.7	1923	4	US-09-252-991A-7983	Sequence 7983, App
C 161	14.2	64.5	810	3	US-09-252-991A-13717	Sequence 27, Appli	234	13.8	62.7	1959	4	US-09-252-991A-8169	Sequence 8169, App
C 162	14.2	64.5	933	4	US-09-252-991A-8451	Sequence 8451, App	235	13.8	62.7	2010	4	US-09-252-991A-8169	Sequence 1, Appli
C 163	14.2	64.5	966	4	US-09-489-039A-2178	Sequence 2178, App	236	13.8	62.7	2230	4	US-08-200-512-1	Sequence 1, Appli
C 164	14.2	64.5	1011	4	US-09-252-991A-16115	Sequence 16115, A	237	13.8	62.7	2436	4	US-09-540-236-492	Sequence 492, App
C 165	14.2	64.5	1089	4	US-09-252-991A-13652	Sequence 13652, A	238	13.8	62.7	2513	4	US-09-585-173B-3	Sequence 3, Appli
C 166	14.2	64.5	1110	4	US-09-252-991A-8388	Sequence 8388, App	239	13.8	62.7	2642	4	US-09-595-002-4	Sequence 4, Appli
C 167	14.2	64.5	1131	4	US-09-252-991A-8486	Sequence 8486, App	240	13.8	62.7	2689	1	US-08-465-795-2	Sequence 2, Appli
C 168	14.2	64.5	1233	4	US-09-543-681A-2246	Sequence 2246, App	241	13.8	62.7	2736	4	US-09-252-991A-8666	Sequence 8666, App
C 169	14.2	64.5	1401	1	US-07-843-949A-3	Sequence 3, Appli	242	13.8	62.7	3245	1	US-07-935-311A-3	Sequence 3, Appli
C 170	14.2	64.5	1401	2	US-08-218-978-3	Sequence 3, Appli	243	13.8	62.7	3245	1	PCT-US93-07996-3	Sequence 3, Appli
C 171	14.2	64.5	1477	4	US-09-976-594-794	Sequence 794, Appli	244	13.8	62.7	4307	1	US-08-190-687B-7	Sequence 7, Appli
C 172	14.2	64.5	1503	4	US-09-252-991A-13684	Sequence 13684, A	245	13.8	62.7	5092	4	US-09-620-312D-153	Sequence 153, App
C 173	14.2	64.5	1644	3	US-08-948-564-5	Sequence 5, Appli	246	13.8	62.7	5494	4	US-09-437-568A-45	Sequence 45, Appli

247 13.8 62.7 5889 1 US-08-465-795-1 Sequence 1, Appl
 C 248 13.8 62.7 5889 4 US-10-204-708-6 Sequence 6, Appl
 C 249 13.8 62.7 7001 1 US-08-258-261B-1 Sequence 1, Appl
 C 250 13.8 62.7 7001 1 US-08-456-837-1 Sequence 1, Appl
 C 251 13.8 62.7 7001 1 US-08-457-342-1 Sequence 1, Appl
 C 252 13.8 62.7 7001 1 US-08-457-646A-1 Sequence 1, Appl
 C 253 13.8 62.7 7001 1 US-08-458-076A-1 Sequence 1, Appl
 C 254 13.8 62.7 7001 1 US-08-457-338A-1 Sequence 1, Appl
 C 255 13.8 62.7 7001 1 US-08-761-258-6 Sequence 6, Appl
 C 256 13.8 62.7 7001 1 US-08-729-214-1 Sequence 1, Appl
 C 257 13.8 62.7 7001 1 US-08-977-306-6 Sequence 6, Appl
 C 258 13.8 62.7 7001 3 US-09-028-934-1 Sequence 14, Appl
 C 259 13.8 62.7 12886 4 US-09-453-702B-14 Sequence 1, Appl
 C 260 13.8 62.7 65792 4 US-09-596-002-31 Sequence 31, Appl
 C 261 13.8 62.7 1230025 4 US-09-198-452A-1 Sequence 1, Appl
 C 262 13.8 62.7 1664976 4 US-08-916-421B-1 Sequence 1, Appl
 C 263 13.6 61.8 35 1 US-08-361-337-20 Sequence 20, Appl
 C 264 13.6 61.8 372 4 US-09-266-965-65 Sequence 65, Appl
 C 265 13.6 61.8 390 4 US-09-134-000C-495 Sequence 495, Appl
 C 266 13.6 61.8 405 4 US-09-489-039A-5545 Sequence 5545, Appl
 C 267 13.6 61.8 537 4 US-09-489-039A-2077 Sequence 2077, Appl
 C 268 13.6 61.8 542 3 US-08-818-112-15 Sequence 15, Appl
 C 269 13.6 61.8 542 4 US-08-818-111-15 Sequence 15, Appl
 C 270 13.6 61.8 542 4 US-09-056-556-15 Sequence 15, Appl
 C 271 13.6 61.8 542 4 US-09-072-596-15 Sequence 15, Appl
 C 272 13.6 61.8 542 4 US-09-072-596-15 Sequence 15, Appl
 C 273 13.6 61.8 600 4 US-09-669-751-245 Sequence 245, Appl
 C 274 13.6 61.8 632 4 US-09-252-991A-6234 Sequence 6234, Appl
 C 275 13.6 61.8 636 4 US-09-429-906B-3 Sequence 3, Appl
 C 276 13.6 61.8 753 4 US-09-252-991A-16180 Sequence 16180, A
 C 277 13.6 61.8 759 4 US-09-252-991A-16011 Sequence 16011, A
 C 278 13.6 61.8 771 4 US-09-489-039A-4829 Sequence 4829, Appl
 C 279 13.6 61.8 783 4 US-09-252-991A-14448 Sequence 14448, A
 C 280 13.6 61.8 786 4 US-09-489-039A-3554 Sequence 3554, Appl
 C 281 13.6 61.8 786 4 US-08-891-641-22 Sequence 22, Appl
 C 282 13.6 61.8 939 4 US-09-252-991A-13394 Sequence 13394, A
 C 283 13.6 61.8 960 4 US-09-328-352-3516 Sequence 3516, Appl
 C 284 13.6 61.8 962 4 US-08-936-165A-140 Sequence 140, Appl
 C 285 13.6 61.8 979 4 US-09-252-991A-13601 Sequence 13601, A
 C 286 13.6 61.8 1008 4 US-09-489-039A-5751 Sequence 5751, Appl
 C 287 13.6 61.8 1017 4 US-09-345-236B-46 Sequence 46, Appl
 C 288 13.6 61.8 1017 4 US-09-345-236B-46 Sequence 46, Appl
 C 289 13.6 61.8 1029 4 US-09-252-991A-6399 Sequence 6399, Appl
 C 290 13.6 61.8 1030 3 US-08-858-003-2 Sequence 2, Appl
 C 291 13.6 61.8 1030 3 US-09-078-166-2 Sequence 2, Appl
 C 292 13.6 61.8 1030 3 US-08-397-467-2 Sequence 2, Appl
 C 293 13.6 61.8 1038 4 US-09-543-681A-2620 Sequence 2620, Appl
 C 294 13.6 61.8 1105 4 US-09-072-596-329 Sequence 329, Appl
 C 295 13.6 61.8 1105 4 US-09-072-596-334 Sequence 334, Appl
 C 296 13.6 61.8 1138 4 US-10-164-595-33 Sequence 33, Appl
 C 297 13.6 61.8 1140 4 US-09-543-681A-3785 Sequence 3785, Appl
 C 298 13.6 61.8 1164 4 US-09-252-991A-6068 Sequence 6068, Appl
 C 299 13.6 61.8 1209 4 US-09-252-991A-137 Sequence 137, Appl
 C 300 13.6 61.8 1239 4 US-09-543-681A-4010 Sequence 4010, Appl

ALIGNMENTS

RESULT 1
 US-09-103-840A-2/c
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 4403765
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; OTHER INFORMATION: CDC 1551
 ; OTHER INFORMATION: "a" bases at various positions throughout the sequence
 ; OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Query Match 100.0%; Score 22; DB 3; Length 4403765;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22
 DB 3234122 CCGGCAAAACCTCGGCTAAC 3234101

RESULT 2

US-09-103-840A-1/c
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 4411529
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; OTHER INFORMATION: H37RV
 US-09-103-840A-1

Query Match 100.0%; Score 22; DB 3; Length 4411529;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22
 DB 3239846 CCGGCAAAACCTCGGCTAAC 3239825

RESULT 3

US-08-680-506-5
 ; Sequence 5, Application US/08680506C
 ; Patent No. 6008013
 ; GENERAL INFORMATION:
 ; APPLICANT: REYNOLDS, Paul R.
 ; TITLE OF INVENTION: CHONDROCYTE PROTEINS
 ; FILE REFERENCE: 176/60091
 ; CURRENT APPLICATION NUMBER: US/08/680,506C
 ; CURRENT FILING DATE: 1996-07-08
 ; EARLIER APPLICATION NUMBER: 60/021,672
 ; EARLIER FILING DATE: 1996-07-05
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 954
 ; TYPE: DNA
 ; ORGANISM: Gallus gallus
 US-08-680-506-5

```
Query Match      71.8%; Score 15.8; DB 3; Length 954;
Best Local Similarity 89.5%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CCGGGCAAAACCTCGGCT 19
Db      227 CCGGGCAAAATCGCGGCT 245
          |||||
          |||||
          |||||
          |||||
          |||||

RESULT 4
US-08-680-506-8
; Sequence 8, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-8

Query Match      71.8%; Score 15.8; DB 3; Length 1347;
Best Local Similarity 89.5%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CCGGGCAAAACCTCGGCT 19
Db      227 CCGGGCAAAATCGCGGCT 245
          |||||
          |||||
          |||||
          |||||
          |||||

RESULT 5
US-08-680-506-6
; Sequence 6, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-6

Query Match      71.8%; Score 15.8; DB 3; Length 1587;
Best Local Similarity 89.5%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CCGGGCAAAACCTCGGCT 19
Db      307 CCGGGCAAAATCGCGGCT 325
          |||||
          |||||
          |||||
          |||||
          |||||

RESULT 6
US-08-680-506-4
; Sequence 4, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2233
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-4

Query Match      71.8%; Score 15.8; DB 3; Length 2233;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CCGGGCAAAACCTCGGCT 19
Db      307 CCGGGCAAAATCGCGGCT 325
          |||||
          |||||
          |||||
          |||||
          |||||

RESULT 7
US-08-680-506-2
; Sequence 2, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5027
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-2

Query Match      71.8%; Score 15.8; DB 3; Length 5027;
Best Local Similarity 89.5%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CCGGGCAAAACCTCGGCT 19
Db      307 CCGGGCAAAATCGCGGCT 325
          |||||
          |||||
          |||||
          |||||
          |||||

RESULT 8
US-09-328-352-1452/c
; Sequence 1452, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1452
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1452

Query Match      70.9%; Score 15.6; DB 4; Length 876;
Best Local Similarity 81.8%; Pred. No. 95;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

5229279-3/c
; Patent No. 5229279
; APPLICANT: PEOPLES, OLIVER P.;SINSKEY,ANTHONY J.
; TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER
; BIOPOLYMERS
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/556,535
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-AUG-1987
; SEQ ID NO:3:
; LENGTH: 2044
5229279-3
Query Match 70.9%; Score 15.6; DB 6; Length 2044;
Best Local Similarity 81.8%; Pred. No. 1.1e-02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTTAC 22
Db 1263 CCGGCAGAACCGCGCTCAC 1242
RESULT 12
5512669-3/c
; Patent No. 5512669
; APPLICANT: PEOPLES, OLIVER P.;SINSKEY,ANTHONY J.
; TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
; REDUCTASE
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,667
; FILING DATE: 29-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 124,570
; FILING DATE: 20-SEP-1993
; APPLICATION NUMBER: 944,488
; FILING DATE: 03-NOV-1992
; APPLICATION NUMBER: 566,535
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-JUN-1987
; SEQ ID NO:3:
; LENGTH: 2094
5512669-3
Query Match 70.9%; Score 15.6; DB 6; Length 2094;
Best Local Similarity 81.8%; Pred. No. 1.1e-02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTTAC 22
Db 1313 CCGGCAGAACCGCGCTCAC 1292
RESULT 13
US-10-204-708-40/c
; Sequence 40, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8

US-09-266-965-38
; Sequence 38, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-38
Query Match 70.9%; Score 15.6; DB 4; Length 1041;
Best Local Similarity 81.8%; Pred. No. 97;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTTAC 22
Db 506 CCGGGAACACATCCGCTCAC 527
RESULT 10
US-09-489-039A-3575
; Sequence 3575, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3575
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3575
Query Match 70.9%; Score 15.6; DB 4; Length 1206;
Best Local Similarity 81.8%; Pred. No. 99;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTTAC 22
Db 631 CTGGCAATATCTCCGCTATC 652
RESULT 11

; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 40
; LENGTH: 19513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-40

Query Match 70.9%; Score 15.6; DB 4; Length 19513;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 40 CGGATCAAAACCTCCACCAAC 19

RESULT 14
US-09-596-002-13/c
; Sequence 13, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 31940
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 13
; PUBLICATION INFORMATION:
US-09-596-002-13

Query Match 70.9%; Score 15.6; DB 4; Length 31940;
Best Local Similarity 81.8%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 30387 CCGGGCAAAACCTCCGCAAC 30366

RESULT 15
US-09-266-965-76/c
; Sequence 76, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965

; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Query Match 70.9%; Score 15.6; DB 4; Length 53500;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 9734 CCGGGCAACACATCCGCTTAC 9713

RESULT 16
US-09-489-039A-2975
; Sequence 2975, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2975
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2975

Query Match 69.1%; Score 15.2; DB 4; Length 273;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CCGGGCAAAACCTCGCCTAA 21
DB 205 CCGGGCAAAACCTCTCCGAA 224

RESULT 17
US-09-489-039A-2954/c
; Sequence 2954, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2954
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2954

Query Match 69.1%; Score 15.2; DB 4; Length 447;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGGCAAAACCTCGGCTAA 21
Db 421 CCGGCAAAACCTCGGCA 402

RESULT 18
US-09-107-532A-2574/c
; Sequence 2574, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Denise
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2574:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (b) LOCATION 1...534
; SEQUENCE DESCRIPTION: SEQ ID NO: 2574:
US-09-107-532A-2574

Query Match 69.1%; Score 15.2; DB 4; Length 534;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CCGCAAAACCTCGGCTAAC 22
Db 247 GGGCAAAACCGACTCTTAC 228

RESULT 19
US-09-976-594-625/c

; Sequence 625, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 625
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1344185CB1
; NAME/KEY: unsure
; LOCATION: 7, 19, 25, 32, 45, 59, 88, 96, 108, 116, 124, 150, 167, 181, 191, 216, 359,
; LOCATION: 221, 226, 233, 246, 264, 276, 286, 303, 307, 315, 330, 332, 336, 351, 359,
; LOCATION: 363, 373, 377, 382, 395, 407, 424, 431, 435, 442, 446, 470, 488, 492, 499,
; LOCATION: 504, 537, 551, 556, 583, 592, 596, 604, 615, 641, 671, 696, 709, 714, 726,
; LOCATION: 739, 747, 751, 756, 782, 774, 810, 842, 844, 906
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-625

Query Match 69.1%; Score 15.2; DB 4; Length 1588;
Best Local Similarity 81.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGCAAAACCTCGGCTAA 21
Db 584 CNGGGCAAAACCCGCCCAAA 564

RESULT 20

US-09-489-039A-2299
; Sequence 2299, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2299
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2299

Query Match 69.1%; Score 15.2; DB 4; Length 1791;
Best Local Similarity 85.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGGCAAAACCTCGGCTAA 21
Db 1700 CCGTAAACCTCGGCGAA 1719

RESULT 21

US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 69.1%; Score 15.2; DB 4; Length 1830121;

Best Local Similarity 85.0%; Pred. No. 2.3e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCTAA 21

Db 430196 CGAGCAAAACCTTGGCTAA 430177

RESULT 22

US-09-643-990A-1/c

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments

Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 69.1%; Score 15.2; DB 4; Length 1830121;

Best Local Similarity 85.0%; Pred. No. 2.3e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCTAA 21

Db 430196 CGAGCAAAACCTTGGCTAA 430177

RESULT 23

US-08-727-034-5

Sequence 5, Application US/08727034

Patent No. 5665872

GENERAL INFORMATION:

APPLICANT: SAITO, YASHUSHI

APPLICANT: IWASAKI, AKIO

APPLICANT: ARAI, KOICHI

APPLICANT: YAMAZAKI, HIROYUKI

TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND

TITLE OF INVENTION: THE GENE CODING THEREFOR

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: P. C.

ADDRESSEE: P. C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/727,034

FILING DATE: 08-OCT-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 261440/1995

FILING DATE: 09-OCT-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 102451/1996

FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 6642 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-727-034-5

Query Match 68.2%; Score 15; DB 1; Length 6642;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGCGAAAACCTCCG 16
Db 5828 CGGCGAAAACCTCCG 5842

RESULT 24
US-08-727-034-6
Sequence 6, Application US/08727034
Patent No. 5665872
GENERAL INFORMATION:
APPLICANT: SAITO, YAGUSHI
APPLICANT: IWASAKI, AKIO
APPLICANT: ARAI, KOICHI
TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
TITLE OF INVENTION: THE GENE CODING THEREFOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS

LOCATION: 81..6725
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: sig peptide
LOCATION: 81..164
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: misc feature
LOCATION: 165..6722
OTHER INFORMATION: /function= "Nucleotides 165-6722
encode the mature peptide"
OTHER INFORMATION: /note= "Identification Method: S"
US-08-727-034-6

Query Match 68.2%; Score 15; DB 1; Length 6843;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGCGAAAACCTCCG 16
Db 5908 CGGCGAAAACCTCCG 5922

RESULT 25
US-08-858-207A-177
Sequence 177, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: NO. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 177:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-177

Query Match 67.3%; Score 14.8; DB 4; Length 774;
Best Local Similarity 88.9%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCAAAACCTCCGCCTAAC 22
|||||
Db 470 GCAAAACCTCCGCCTGAC 487

RESULT 26

US-09-328-352-715/c
; Sequence 715, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 715
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-715

Query Match 67.3%; Score 14.8; DB 4; Length 1038;
Best Local Similarity 88.9%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
|||||
Db 366 GCGAAACCTCCGCCTAA 349

RESULT 27

US-09-489-039A-1114/c
; Sequence 1114, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1114
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1114

Query Match 67.3%; Score 14.8; DB 4; Length 1206;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCCGCCTA 20
|||||
Db 577 GGGCAAAACCTCCGCCTA 560

RESULT 28

US-09-543-681A-445/c
; Sequence 445, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 445
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-445

Query Match 67.3%; Score 14.8; DB 4; Length 1482;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
|||||
Db 105 GCGTAACCTCCGCCTAA 88

RESULT 29

US-09-620-312D-723/c
; Sequence 723, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, fuding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 723
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (127)..(1401)
US-09-620-312D-723

Query Match 67.3%; Score 14.8; DB 4; Length 1660;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
|||||
Db 363 GCGTAACCTCCGCCTAA 346

RESULT 30

US-09-489-039A-1224
; Sequence 1224, Application US/09489039A
; Patent No. 6610836

; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1224
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1224

Query Match 67.3%; Score 14.8; DB 4; Length 1770;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1194 GGGCAAAACCAACCCGCCGA 1211

Search completed: June 20, 2004, 11:46:26
Job time : 40.453 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 124.434 Seconds
(without alignments)

809.940 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 ccgggcaaaacctccgctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/PCTUS_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	95.5	723	9	US-09-712-363-114
2	17.8	80.9	2629	16	US-10-348-052-24
3	17.2	78.2	1064	13	US-10-424-599-66751
4	17.2	78.2	5925	15	US-10-311-455-1549
5	16.8	76.4	416	13	US-10-424-599-31632
6	16.8	76.4	536	11	US-09-864-408A-265
7	16.8	76.4	1728	16	US-10-260-238-1137
8	16.4	74.5	927	9	US-09-815-242-7429
9	16.4	74.5	933	13	US-10-335-977-3347
10	16.4	74.5	1035	13	US-10-335-977-3348
11	16.4	74.5	1320	13	US-10-425-114-6455
12	16.4	74.5	1320	13	US-10-282-122A-36715
13	16.4	74.5	1385	13	US-10-425-114-29496
14	16.4	74.5	1492	13	US-10-424-599-102731

15	16.4	74.5	3832	15	US-10-102-524-1718	Sequence 1718, Ap
16	16.4	74.5	3832	15	US-10-376-564-86	Sequence 86, Appl
17	16.4	74.5	3832	16	US-10-159-563-291	Sequence 291, Appl
18	16.2	73.6	734	13	US-10-027-632-144898	Sequence 144898,
19	16.2	73.6	734	13	US-10-027-632-144898	Sequence 144898,
20	16.2	73.6	753	12	US-09-727-892-10	Sequence 10, Appl
21	16.2	73.6	1911	12	US-09-973-674A-33	Sequence 33, Appl
22	16.2	73.6	6022	13	US-10-221-714A-119	Sequence 119, Appl
23	16.2	73.6	6022	15	US-10-239-676-79	Sequence 79, Appl
24	16.2	73.6	6022	15	US-10-311-455-1045	Sequence 1045, Ap
25	16.2	73.6	6022	15	US-10-240-455-85	Sequence 85, Appl
26	16.2	73.6	9888	15	US-10-311-455-1214	Sequence 1214, Ap
27	16.2	73.6	16668	12	US-09-727-892-1	Sequence 1, Appli
28	15.8	71.8	158	13	US-10-065-783A-15488	Sequence 15488, A
29	15.8	71.8	158	16	US-10-242-535A-15488	Sequence 15488, A
30	15.8	71.8	346	15	US-10-106-698-1566	Sequence 1566, Ap
31	15.8	71.8	611	13	US-10-424-599-31840	Sequence 31840, A
32	15.8	71.8	990	16	US-10-360-238-4066	Sequence 4066, Ap
33	15.8	71.8	1019	13	US-10-425-114-27844	Sequence 27844, A
34	15.8	71.8	1101	9	US-09-972-186A-3	Sequence 3, Appli
35	15.8	71.8	1375	13	US-10-183-687-466	Sequence 466, App
36	15.8	71.8	1375	15	US-10-180-375-213	Sequence 213, App
37	15.8	71.8	1403	16	US-10-369-493-38599	Sequence 38599, A
38	15.8	71.8	1415	16	US-10-369-493-35378	Sequence 35378, A
39	15.8	71.8	1415	16	US-10-369-493-38195	Sequence 38195, A
40	15.8	71.8	2307	13	US-10-282-122A-12392	Sequence 12392, A
41	15.8	71.8	2507	15	US-10-085-959-53	Sequence 53, Appl
42	15.6	70.9	117	9	US-09-867-701-8023	Sequence 8023, Ap
43	15.6	70.9	376	13	US-10-424-599-52390	Sequence 52390, A
44	15.6	70.9	403	9	US-09-960-352-3319	Sequence 3319, Ap
45	15.6	70.9	473	16	US-10-260-238-3779	Sequence 3779, Ap
46	15.6	70.9	558	13	US-10-282-122A-40260	Sequence 40260, A
47	15.6	70.9	957	13	US-10-282-122A-26820	Sequence 26820, A
48	15.6	70.9	1041	10	US-09-953-348-38	Sequence 38, Appl
49	15.6	70.9	1041	15	US-10-267-255-38	Sequence 38, Appl
50	15.6	70.9	1150	16	US-10-260-238-3780	Sequence 3780, Ap
51	15.6	70.9	1182	13	US-10-282-122A-23981	Sequence 23981, A
52	15.6	70.9	1417	17	US-10-426-002-21	Sequence 21, Appl
53	15.6	70.9	1915	13	US-10-027-632-98082	Sequence 98082, A
54	15.6	70.9	1915	13	US-10-027-632-98325	Sequence 98325, A
55	15.6	70.9	1915	16	US-10-027-632-98082	Sequence 98082, A
56	15.6	70.9	2400	16	US-10-369-493-37751	Sequence 37751, A
57	15.6	70.9	2497	13	US-10-424-599-14892	Sequence 14892, A
58	15.6	70.9	5663	15	US-10-311-455-670	Sequence 670, App
59	15.6	70.9	5663	17	US-10-240-589C-32	Sequence 32, Appl
60	15.6	70.9	7029	9	US-09-826-191-1	Sequence 1, Appli
61	15.6	70.9	7029	15	US-10-263-666-1	Sequence 1, Appli
62	15.6	70.9	7029	15	US-10-261-481-1	Sequence 1, Appli
63	15.6	70.9	7029	15	US-10-261-942-1	Sequence 1, Appli
64	15.6	70.9	8115	9	US-09-778-516A-1	Sequence 1, Appli
65	15.6	70.9	8115	9	US-09-778-516A-2	Sequence 2, Appli
66	15.6	70.9	10183	13	US-10-221-714A-478	Sequence 478, App
67	15.6	70.9	19513	13	US-10-221-613-206	Sequence 206, App
68	15.6	70.9	19513	15	US-10-284-708-40	Sequence 40, Appl
69	15.6	70.9	31940	13	US-10-672-787-13	Sequence 13, Appl
70	15.6	70.9	43411	17	US-10-450-826-76	Sequence 76, Appl
71	15.6	70.9	53500	10	US-09-953-348-76	Sequence 76, Appl
72	15.6	70.9	53500	15	US-10-267-255-76	Sequence 106, App
73	15.6	70.9	96593	12	US-10-052-482-106	Sequence 43, Appl
74	15.6	70.9	514	9	US-09-833-790-43	Sequence 27418, A
75	15.4	70.0	514	13	US-10-027-632-274118	Sequence 274118,
76	15.4	70.0	548	13	US-10-027-632-274118	Sequence 45, Appl
77	15.4	70.0	548	16	US-09-811-284-45	Sequence 105614,
78	15.4	70.0	616	9	US-10-027-632-105614	Sequence 105614,
79	15.4	70.0	674	13	US-10-027-632-105614	Sequence 101014,
80	15.4	70.0	674	13	US-10-027-632-101014	Sequence 101014,
81	15.4	70.0	1066	13	US-10-369-493-26789	Sequence 26789, A
82	15.4	70.0	1066	16	US-10-369-493-26789	Sequence 26789, A
83	15.4	70.0	1149	16	US-09-974-300-2965	Sequence 5924, Ap
84	15.4	70.0	1233	9	US-09-815-242-5924	Sequence 5924, Ap
85	15.4	70.0	1401	9	US-10-282-122A-20221	Sequence 20221, A
86	15.4	70.0	1401	13	US-10-425-114-12027	Sequence 12027, A
87	15.4	70.0	2059	13	US-10-425-114-12027	Sequence 12027, A

C 88	15.4	70.0	2485	13	US-10-424-599-105334	Sequence 105334, A	161	14.8	67.3	813	13	US-10-627-476-349	Sequence 349, App
C 89	15.4	70.0	2485	13	US-10-425-114-29802	Sequence 29802, A	162	14.8	67.3	852	13	US-10-425-114-15743	Sequence 15743, A
C 90	15.4	70.0	2616	16	US-10-369-493-42226	Sequence 42226, A	163	14.8	67.3	855	10	US-09-769-787-248	Sequence 248, App
C 91	15.4	70.0	6694	15	US-10-311-455-1637	Sequence 1637, App	164	14.8	67.3	867	13	US-10-282-122A-20463	Sequence 20463, A
C 92	15.4	70.0	8201	15	US-10-240-485-653	Sequence 653, App	165	14.8	67.3	882	9	US-09-974-300-697	Sequence 697, App
C 93	15.4	70.0	12460	15	US-10-311-455-41	Sequence 41, App	166	14.8	67.3	900	13	US-10-027-632-153958	Sequence 153958, A
C 94	15.4	70.0	36595	16	US-10-085-117-10	Sequence 10, App	167	14.8	67.3	900	16	US-10-027-632-153958	Sequence 153958, A
C 95	15.4	70.0	304905	13	US-10-271-416-1	Sequence 1, App	168	14.8	67.3	918	9	US-09-738-626-1054	Sequence 1054, App
C 96	15.2	69.1	510	9	US-09-783-590-3131	Sequence 3131, App	169	14.8	67.3	949	13	US-10-027-632-163668	Sequence 163668, A
C 97	15.2	69.1	876	16	US-10-369-493-32911	Sequence 32911, A	170	14.8	67.3	949	13	US-10-027-632-163668	Sequence 163668, A
C 98	15.2	69.1	876	16	US-10-369-493-33069	Sequence 33069, A	171	14.8	67.3	949	16	US-10-027-632-163668	Sequence 163668, A
C 99	15.2	69.1	895	13	US-10-424-599-1492	Sequence 1492, App	172	14.8	67.3	949	16	US-10-027-632-163668	Sequence 163668, A
C 100	15.2	69.1	915	13	US-10-424-599-84426	Sequence 84426, A	173	14.8	67.3	1047	13	US-10-627-476-347	Sequence 347, App
C 101	15.2	69.1	960	16	US-10-369-493-41172	Sequence 41172, A	174	14.8	67.3	1077	17	US-10-027-632-153958	Sequence 153958, A
C 102	15.2	69.1	996	15	US-10-398-846-6830	Sequence 6830, App	175	14.8	67.3	1077	17	US-10-027-632-153958	Sequence 153958, A
C 103	15.2	69.1	1031	13	US-10-424-599-111658	Sequence 111658, A	176	14.8	67.3	1143	9	US-09-972-186A-1	Sequence 1, App
C 104	15.2	69.1	1031	13	US-10-412-699B-1707	Sequence 1707, App	177	14.8	67.3	1143	9	US-09-972-186A-1	Sequence 1, App
C 105	15.2	69.1	1031	16	US-10-374-780A-1864	Sequence 1864, App	178	14.8	67.3	1218	9	US-09-815-242-9704	Sequence 9704, App
C 106	15.2	69.1	1050	13	US-10-282-122A-28092	Sequence 28092, A	179	14.8	67.3	1311	13	US-10-282-122A-28092	Sequence 28092, A
C 107	15.2	69.1	1230	13	US-10-282-122A-28865	Sequence 28865, A	180	14.8	67.3	1311	13	US-10-282-122A-28865	Sequence 28865, A
C 108	15.2	69.1	1254	13	US-10-282-122A-29724	Sequence 29724, A	181	14.8	67.3	1359	13	US-10-282-122A-29724	Sequence 29724, A
C 109	15.2	69.1	1403	13	US-10-425-114-9654	Sequence 9654, App	182	14.8	67.3	1371	17	US-10-275-026A-133	Sequence 133, App
C 110	15.2	69.1	1428	13	US-10-282-122A-21975	Sequence 21975, A	183	14.8	67.3	1371	17	US-10-275-026A-133	Sequence 133, App
C 111	15.2	69.1	1574	16	US-10-431-410-29	Sequence 29, App	184	14.8	67.3	1400	13	US-10-424-599-125443	Sequence 125443, A
C 112	15.2	69.1	1642	13	US-10-425-114-12815	Sequence 12815, A	185	14.8	67.3	1401	9	US-09-815-242-9919	Sequence 9919, App
C 113	15.2	69.1	1901	13	US-10-424-599-8331	Sequence 8331, App	186	14.8	67.3	1401	13	US-10-282-122A-39824	Sequence 39824, A
C 114	15.2	69.1	2000	9	US-09-887-576-831	Sequence 831, App	187	14.8	67.3	1449	13	US-10-282-122A-39824	Sequence 39824, A
C 115	15.2	69.1	2452	9	US-09-410-194-18	Sequence 18, App	188	14.8	67.3	1449	13	US-10-282-122A-39824	Sequence 39824, A
C 116	15.2	69.1	2770	14	US-10-005-921-1	Sequence 1, App	189	14.8	67.3	1483	15	US-10-220-564-1	Sequence 1, App
C 117	15.2	69.1	3741	13	US-10-282-122A-38700	Sequence 38700, A	190	14.8	67.3	1660	15	US-10-037-270-723	Sequence 723, App
C 118	15.2	69.1	4395	13	US-10-282-122A-36344	Sequence 36344, A	191	14.8	67.3	1660	15	US-10-037-270-723	Sequence 723, App
C 119	15.2	69.1	5857	15	US-10-311-455-501	Sequence 501, App	192	14.8	67.3	1660	16	US-10-117-722-723	Sequence 723, App
C 120	15.2	69.1	47884	16	US-10-085-117-121	Sequence 121, App	193	14.8	67.3	1885	13	US-10-282-122A-23378	Sequence 23378, A
C 121	15.2	69.1	1830121	15	US-10-328-960-1	Sequence 1, App	194	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 122	15.2	69.1	1830121	15	US-10-328-960-1	Sequence 1, App	195	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 123	15.2	69.1	9035608	15	US-10-156-761-1	Sequence 1, App	196	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 124	15	68.2	514	13	US-10-424-599-85692	Sequence 85692, A	197	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 125	15	68.2	1440	13	US-10-282-122A-30872	Sequence 30872, A	198	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 126	15	68.2	6645	17	US-10-464-368-44	Sequence 44, App	199	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 127	15	68.2	6840	17	US-10-343-710-72	Sequence 72, App	200	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 128	15	68.2	6868	10	US-09-319-039-39	Sequence 39, App	201	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 129	15	68.2	6868	17	US-10-188-832-77	Sequence 77, App	202	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 130	15	68.2	6906	17	US-10-188-832-77	Sequence 77, App	203	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 131	15	68.2	6981	10	US-09-360-706-977	Sequence 977, App	204	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 132	15	68.2	6981	10	US-09-360-706-977	Sequence 977, App	205	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 133	15	68.2	6981	15	US-10-097-340-299	Sequence 299, App	206	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 134	15	68.2	7460	15	US-10-176-847-93	Sequence 93, App	207	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 135	14.8	67.3	38	10	US-09-877-478-3522	Sequence 3522, App	208	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 136	14.8	67.3	38	10	US-10-342-902-3522	Sequence 3522, App	209	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 137	14.8	67.3	38	15	US-10-156-306-1999	Sequence 1999, App	210	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 138	14.8	67.3	220	9	US-09-764-860-924	Sequence 924, App	211	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 139	14.8	67.3	220	15	US-10-074-095-924	Sequence 924, App	212	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 140	14.8	67.3	220	16	US-10-212-872-924	Sequence 924, App	213	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 141	14.8	67.3	223	13	US-10-424-599-34914	Sequence 34914, App	214	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 142	14.8	67.3	289	13	US-10-424-599-3398	Sequence 3398, App	215	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 143	14.8	67.3	320	13	US-10-424-599-45274	Sequence 45274, A	216	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 144	14.8	67.3	338	13	US-10-424-599-17804	Sequence 17804, A	217	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 145	14.8	67.3	402	10	US-09-918-995-4211	Sequence 4211, App	218	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 146	14.8	67.3	404	13	US-10-424-599-18495	Sequence 18495, App	219	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 147	14.8	67.3	412	9	US-09-728-446-899	Sequence 899, App	220	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 148	14.8	67.3	597	13	US-10-282-122A-29006	Sequence 29006, A	221	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 149	14.8	67.3	597	13	US-10-027-632-49511	Sequence 49511, A	222	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 150	14.8	67.3	597	13	US-10-027-632-49512	Sequence 49512, A	223	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 151	14.8	67.3	597	13	US-10-027-632-49512	Sequence 49512, A	224	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 152	14.8	67.3	597	13	US-10-027-632-83436	Sequence 83436, A	225	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 153	14.8	67.3	597	13	US-10-027-632-83437	Sequence 83437, A	226	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 154	14.8	67.3	597	16	US-10-027-632-49511	Sequence 49511, A	227	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 155	14.8	67.3	597	16	US-10-027-632-49512	Sequence 49512, A	228	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 156	14.8	67.3	597	16	US-10-027-632-83435	Sequence 83435, A	229	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 157	14.8	67.3	597	16	US-10-027-632-83436	Sequence 83436, A	230	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 158	14.8	67.3	597	16	US-10-027-632-83437	Sequence 83437, A	231	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 159	14.8	67.3	690	16	US-10-369-493-32609	Sequence 32609, A	232	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A
C 160	14.8	67.3	697	16	US-10-260-238-3781	Sequence 3781, App	233	14.8	67.3	2028	9	US-10-424-599-46890	Sequence 46890, A

C 234 14.6 66.4 497 10 US-09-918-995-179 Sequence 179, App
C 235 14.6 66.4 500 10 US-09-918-995-32500 Sequence 32500, A
C 236 14.6 66.4 500 10 US-10-085-783A-47225 Sequence 47225, A
C 237 14.6 66.4 500 10 US-10-242-535A-47225 Sequence 47225, A
C 238 14.6 66.4 515 10 US-09-918-995-25778 Sequence 25778, A
C 239 14.6 66.4 551 13 US-10-085-783A-18289 Sequence 18289, A
C 240 14.6 66.4 551 16 US-10-242-535A-18289 Sequence 18289, A
C 241 14.6 66.4 567 9 US-09-796-692-4489 Sequence 4489, App
C 242 14.6 66.4 567 15 US-10-040-862-4489 Sequence 4489, App
C 243 14.6 66.4 567 16 US-10-175-499-61 Sequence 61, Appl
C 244 14.6 66.4 567 16 US-10-057-475B-4489 Sequence 4489, App
C 245 14.6 66.4 567 16 US-10-154-844B-4489 Sequence 4489, App
C 246 14.6 66.4 573 13 US-10-282-122A-19320 Sequence 19320, A
C 247 14.6 66.4 588 13 US-10-027-632-266933 Sequence 266933, A
C 248 14.6 66.4 588 16 US-10-027-632-266933 Sequence 266933, A
C 249 14.6 66.4 594 13 US-10-027-632-29806 Sequence 29806, A
C 250 14.6 66.4 594 16 US-10-027-632-29806 Sequence 29806, A
C 251 14.6 66.4 601 15 US-10-066-543-1384 Sequence 1384, App
C 252 14.6 66.4 616 13 US-10-085-783A-14952 Sequence 14952, A
C 253 14.6 66.4 616 16 US-10-242-535A-14952 Sequence 14952, A
C 254 14.6 66.4 634 13 US-10-425-114-1907 Sequence 1907, App
C 255 14.6 66.4 634 15 US-10-198-846-12781 Sequence 12781, A
C 256 14.6 66.4 644 13 US-10-027-632-243647 Sequence 243647, A
C 257 14.6 66.4 644 16 US-10-027-632-243647 Sequence 243647, A
C 258 14.6 66.4 674 16 US-10-264-049-1663 Sequence 1663, App
C 259 14.6 66.4 682 13 US-10-425-114-21935 Sequence 21935, A
C 260 14.6 66.4 707 15 US-10-231-417-2535 Sequence 25, Appl
C 261 14.6 66.4 730 13 US-10-027-632-26574 Sequence 26574, A
C 262 14.6 66.4 730 16 US-10-027-632-26574 Sequence 26574, A
C 263 14.6 66.4 788 13 US-10-027-632-28591 Sequence 28591, A
C 264 14.6 66.4 788 16 US-10-027-632-28591 Sequence 28591, A
C 265 14.6 66.4 822 16 US-10-369-493-40993 Sequence 40993, A
C 266 14.6 66.4 825 13 US-10-282-122A-28880 Sequence 28880, A
C 267 14.6 66.4 825 16 US-10-282-122A-29744 Sequence 29744, A
C 268 14.6 66.4 845 13 US-10-027-632-126744 Sequence 126744, A
C 269 14.6 66.4 845 16 US-10-027-632-126744 Sequence 126744, A
C 270 14.6 66.4 867 16 US-10-175-499-60 Sequence 60, Appl
C 271 14.6 66.4 872 13 US-10-027-632-122263 Sequence 122263, A
C 272 14.6 66.4 972 13 US-10-027-632-122264 Sequence 122264, A
C 273 14.6 66.4 972 16 US-10-027-632-122264 Sequence 122264, A
C 274 14.6 66.4 972 16 US-10-027-632-122264 Sequence 122264, A
C 275 14.6 66.4 972 16 US-10-084-846A-77 Sequence 77, Appl
C 276 14.6 66.4 1032 16 US-10-330-051A-35 Sequence 35, Appl
C 277 14.6 66.4 1050 13 US-10-156-761-2063 Sequence 2063, App
C 278 14.6 66.4 1206 13 US-10-282-122A-26589 Sequence 26589, A
C 279 14.6 66.4 1209 9 US-09-712-363-1 Sequence 1, Appl
C 280 14.6 66.4 1209 13 US-10-282-122A-28142 Sequence 28142, A
C 281 14.6 66.4 1316 13 US-10-403-571-107 Sequence 107, App
C 282 14.6 66.4 1338 13 US-10-225-066A-43 Sequence 43, Appl
C 283 14.6 66.4 1338 16 US-10-374-780A-205 Sequence 205, App
C 284 14.6 66.4 1349 13 US-10-425-114-32492 Sequence 32492, A
C 285 14.6 66.4 1425 9 US-09-738-626-700 Sequence 700, App
C 286 14.6 66.4 1451 15 US-10-106-698-401 Sequence 401, App
C 287 14.6 66.4 1452 9 US-09-925-299-183 Sequence 183, App
C 288 14.6 66.4 1452 10 US-09-925-299-183 Sequence 183, App
C 289 14.6 66.4 1520 10 US-09-919-039-342 Sequence 342, App
C 290 14.6 66.4 1542 10 US-09-919-039-341 Sequence 341, App
C 291 14.6 66.4 1550 16 US-10-116-275-278 Sequence 278, App
C 292 14.6 66.4 1559 15 US-10-288-930-57 Sequence 57, Appl
C 293 14.6 66.4 1590 9 US-09-738-626-688 Sequence 688, App
C 294 14.6 66.4 1596 13 US-10-282-122A-30755 Sequence 30755, A
C 295 14.6 66.4 1635 13 US-10-363-616-107 Sequence 107, App
C 296 14.6 66.4 1691 13 US-10-198-846-13424 Sequence 13424, A
C 297 14.6 66.4 1819 16 US-10-062-674-1807 Sequence 1807, App
C 298 14.6 66.4 1846 15 US-10-197-666A-77 Sequence 77, Appl
C 299 14.6 66.4 1892 15 US-10-252-157-41 Sequence 41, Appl
C 300 14.6 66.4 1896 16 US-10-108-260A-2372 Sequence 2372, App

ALIGNMENTS

US-09-712-363-114
; Sequence 114, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/136,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-114
Query Match 95.5%; Score 21; DB 9; Length 723;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CCGGCAAAACCTCGCCTAA 21
DB 703 CCGGCAAAACCTCGCCTAA 723
RESULT 2
US-10-348-052-24/c
; Sequence 24, Application US/10348052
; Publication No. US20030219782A1
; GENERAL INFORMATION:
; APPLICANT: Saba, Julie D.
; APPLICANT: Fyrist, Henrik
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION
; OF SPHINGOLIPID METABOLISM AND/OR SIGNALING
; FILE REFERENCE: 200116.405
; CURRENT APPLICATION NUMBER: US/10/348,052
; CURRENT FILING DATE: 2003-01-17
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-348-052-24
Query Match 80.9%; Score 17.8; DB 16; Length 2629;
Best Local Similarity 90.5%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 CCGGCAAAACCTCGCCTAA 21

APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rick, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 1137
LENGTH: 1728
TYPE: DNA
ORGANISM: Oryza sativa
US-10-260-238-1137

Query Match 76.4%; Score 16.8; DB 16; Length 1728;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGGCTAAC 22
DB 337 GGGCAAAACCTCGGCTAAC 356

RESULT 8
US-09-815-242-7429/c
Sequence 7429, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Chlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7429
LENGTH: 927
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)....(927)
US-09-815-242-7429

Query Match 74.5%; Score 16.4; DB 9; Length 927;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCGGCTAA 21
DB 249 GGCAAAACCTCGGCTAA 232

RESULT 9
US-10-335-977-3347/c
Sequence 3347, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3347:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...927
SEQUENCE DESCRIPTION: SEQ ID NO: 3347:
US-10-335-977-3347

Query Match 74.5%; Score 16.4; DB 13; Length 927;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCGGCTAA 21
DB 249 GGCAAAACCTCGGCTAA 232

RESULT 10
US-10-335-977-3348/c
Sequence 3348, Application US/10335977

Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3348:
SEQUENCE CHARACTERISTICS:
LENGTH: 933 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...933
SEQUENCE DESCRIPTION: SEQ ID NO: 3348:

US-10-335-977-3348
Query Match 74.5%; Score 16.4; DB 13; Length 933;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GGCAAAACCTCCGCCTAA 21
Db 258 GGCAAAACCTCCGCCTAA 241

RESULT 11

US-10-425-114-6455
Sequence 6455, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:

APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114

CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 6455
LENGTH: 1035
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 700565645_FLI
US-10-425-114-6455

Query Match 74.5%; Score 16.4; DB 13; Length 1035;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GGCAAAACCTCCGCCTAA 21
Db 99 GGCAAAACCTCCGCCTAA 116

RESULT 12

US-10-282-122A-36715/C
Sequence 36715, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1

SEQ ID NO 36715
LENGTH: 1320
TYPE: DNA

ORGANISM: Salmonella paratyphi A
US-10-282-122A-36715

Query Match 74.5%; Score 16.4; DB 13; Length 1320;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCCGCC 18
Db 1237 CCAGGCAAAACCTCCGCC 1220

RESULT 13
US-10-425-114-29496
; Sequence 29496, Application US/10425114
; Publication No. US20040034988A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 71128
; SEQ ID NO 29496
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMRONOIR024E05_FLI
US-10-425-114-29496

Query Match 74.5%; Score 16.4; DB 13; Length 1385;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCCGCTAA 21
Db 335 GGCAAAACCTCCCTCTAA 352

RESULT 14
US-10-424-599-102731
; Sequence 102731, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 102731
; LENGTH: 1492
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1492)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63784C.1
US-10-424-599-102731

Query Match 74.5%; Score 16.4; DB 13; Length 1492;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCCGCTAA 21
Db 353 GGCAAAACCTCCCTCTAA 370

RESULT 15
US-10-102-524-1718
; Sequence 1718, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1718
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1718

Query Match 74.5%; Score 16.4; DB 15; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCT 19
Db 1164 CGGGCAAAACCTCCACCT 1181

RESULT 16
US-10-376-564-86
; Sequence 86, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-29
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-376-564-86

Query Match 74.5%; Score 16.4; DB 15; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCT 19
Db 1164 CGGGCAAAACCTCCACCT 1181

```
RESULT 17
US-10-159-563-291
; Sequence 291, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Metzger, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 291
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-291

Query Match 74.5%; Score 16.4; DB 16; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCT 19
DB 1164 CGGGCAAAACCTCCACCT 1181

RESULT 18
US-10-027-632-144898/c
; Sequence 144898, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144898
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144898

Query Match 73.6%; Score 16.2; DB 13; Length 734;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGCAAAACCTCCGCCTAA 21
DB 1164 CGGGCAAAACCTCCACCT 1181

RESULT 19
US-10-027-632-144898/c
; Sequence 144898, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144898
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144898

Query Match 73.6%; Score 16.2; DB 16; Length 734;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGCAAAACCTCCGCCTAA 21
DB 515 CTGGCAAAACCTAAGCCTAA 495

RESULT 20
US-09-727-892-10/c
; Sequence 10, Application US/09727892
; Publication No. US20040091856A1
; GENERAL INFORMATION:
; APPLICANT: Phagotech, Inc.
; APPLICANT: BELLETIER, Jerry
; APPLICANT: GROS, Philippe
; APPLICANT: DUBOW, Michael
; TITLE OF INVENTION: DATA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44 AHJD
; TITLE OF INVENTION: THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES
; FILE REFERENCE: 073406-0302
; CURRENT APPLICATION NUMBER: US/09/727,892
; CURRENT FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 10
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Staphylococcus aureus Bacteriophage 44 AHJD
US-09-727-892-10

Query Match 73.6%; Score 16.2; DB 12; Length 753;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCTAAC 22
DB 515 CTGGCAAAACCTAAGCCTAA 495
```

Db 344 CGCTCAAAACCGCGCTTAC 324

RESULT 21

US-09-973-674A-33

; Sequence 33, Application US/09973674A

; Publication No. US20040086937A1

; GENERAL INFORMATION:

; APPLICANT: Lofrer, H.

; APPLICANT: Jacobi, A.

; TITLE OF INVENTION: NOVEL METHOD FOR IDENTIFYING ANTIBACTERIAL COMPOUNDS

; FILE REFERENCE: GPCG-P01-019

; CURRENT APPLICATION NUMBER: US/09/973,674A

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: PCT/EP00/03135

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: EP 00102111.2

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: EP 95107031.9

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 33

; LENGTH: 1911

; TYPE: DNA

; ORGANISM: Escherichia coli

US-09-973-674A-33

Query Match 73.6%; Score 16.2; DB 12; Length 1911;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTTAC 22

Db 146 CGGCAAAACCTACGCTTACC 166

RESULT 22

US-10-221-714A-119/c

; Sequence 119, Application US/10221714A

; Publication No. US20040048254A1

; GENERAL INFORMATION:

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: OLEK, Alexander

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with

; FILE REFERENCE: 5013.1005

; CURRENT APPLICATION NUMBER: US/10/221,714A

; PRIOR FILING DATE: 2003-01-21

; PRIOR APPLICATION NUMBER: PCT/EP01/02955

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: DE 10013947.0

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: DE 10019058.8

; PRIOR FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: DE 10019173.8

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: DE 10039529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 540

; SEQ ID NO 119

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-221-714A-119

Query Match 73.6%; Score 16.2; DB 13; Length 6022;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTTAC 22

Db 5377 CGGCAAAACGTCCGCCGAC 5357

RESULT 23

US-10-239-676-79/c

; Sequence 79, Application US/10239676

; Publication No. US20030082609A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

; FILE REFERENCE: 5013.1003

; CURRENT APPLICATION NUMBER: US/10/239,676

; PRIOR FILING DATE: 2002-09-24

; PRIOR APPLICATION NUMBER: PCT/EP01/03968

; DE 10019058.8

; DE 10019173.8

; DE 10032529.7

; DE 10043826.1

; PRIOR FILING DATE: 2001-04-06

; 2000-04-06

; 2000-04-07

; 2000-06-30

; NUMBER OF SEQ ID NOS: 228

; SEQ ID NO 79

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-79

Query Match 73.6%; Score 16.2; DB 15; Length 6022;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTTAC 22

Db 5377 CGGCAAAACGTCCGCCGAC 5357

RESULT 24

US-10-311-455-1045/c

; Sequence 1045, Application US/10311455

; Publication No. US20030143606A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining

; FILE REFERENCE: 5013.1014

; CURRENT APPLICATION NUMBER: US/10/311,455

; PRIOR FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537

; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 2424

; SEQ ID NO 1045

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

<p>US-10-311-455-1045</p> <p>Query Match 73.6%; Score 16.2; DB 15; Length 6022; Best Local Similarity 85.7%; Pred. No. 3.6e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p> <p>QY 2 CGGGCAAAACCTCCGCTTAAAC 22 DB 5377 CGCGCAAAACGTCCGCCGAAC 5357 </p> <p>RESULT 25</p> <p>US-10-240-453-85/c</p> <p>; Sequence 85, Application US/10240453 ; Publication No. US20030148326A1 ; GENERAL INFORMATION: ; APPLICANT: OLEK, Alexander ; APPLICANT: PIEPENBROCK, Christian ; APPLICANT: BERLIN, Kurt ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA ; TITLE OF INVENTION: Transcription ; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated ; TITLE OF INVENTION: With DNA Transcription ; FILE REFERENCE: 5013.1009 ; CURRENT APPLICATION NUMBER: US/10/240,453 ; PRIOR FILING DATE: 2002-10-02 ; PRIOR APPLICATION NUMBER: PCT/EP01/03973 ; PRIOR FILING DATE: 2001-04-06 ; PRIOR APPLICATION NUMBER: DE 10019058.8 ; PRIOR FILING DATE: 2000-04-06 ; PRIOR APPLICATION NUMBER: DE 10019173.8 ; PRIOR FILING DATE: 2000-04-07 ; PRIOR APPLICATION NUMBER: DE 10032529.7 ; PRIOR FILING DATE: 2000-06-30 ; PRIOR APPLICATION NUMBER: DE 10043826.1 ; PRIOR FILING DATE: 2000-09-01 ; NUMBER OF SEQ ID NOS: 350 ; SEQ ID NO 85 ; LENGTH: 6022 ; TYPE: DNA ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)</p> <p>US-10-240-453-85</p>	<p>US-10-311-455-1214</p> <p>Query Match 73.6%; Score 16.2; DB 15; Length 6022; Best Local Similarity 85.7%; Pred. No. 3.6e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p> <p>QY 2 CGGGCAAAACCTCCGCTTAAAC 22 DB 5377 CGCGCAAAACGTCCGCCGAAC 5357 </p> <p>RESULT 26</p> <p>US-10-311-455-1214/c</p> <p>; Sequence 1214, Application US/10311455 ; Publication No. US20030143606A1 ; GENERAL INFORMATION: ; APPLICANT: OLEK, Alexander ; APPLICANT: PIEPENBROCK, Christian ; APPLICANT: BERLIN, Kurt ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation ; TITLE OF INVENTION: Cytosine methylation ; FILE REFERENCE: 5013.1014 ; CURRENT APPLICATION NUMBER: US/10/311,455 ; PRIOR FILING DATE: 2002-12-16 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537 ; PRIOR FILING DATE: 2001-07-02 ; PRIOR APPLICATION NUMBER: DE 10032529.7 ; PRIOR FILING DATE: 2000-06-30 ; PRIOR APPLICATION NUMBER: DE 10043826.1 ; PRIOR FILING DATE: 2000-09-01</p>	<p>NUMBER OF SEQ ID NOS: 2424 SEQ ID NO 1214 LENGTH: 9888 TYPE: DNA ORGANISM: Artificial Sequence FEATURE: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)</p> <p>US-10-311-455-1214</p> <p>Query Match 73.6%; Score 16.2; DB 15; Length 9888; Best Local Similarity 85.7%; Pred. No. 3.6e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p> <p>QY 2 CGGGCAAAACCTCCGCTTAAAC 22 DB 5084 CGGCCAAAACCTCCCCCTAAC 5064 </p> <p>RESULT 27</p> <p>US-09-727-892-1/c</p> <p>; Sequence 1, Application US/09727892 ; Publication No. US20040091856A1 ; GENERAL INFORMATION: ; APPLICANT: Phagotech, Inc. ; APPLICANT: PELLETIER, Jerry ; APPLICANT: GROS, Philippe ; APPLICANT: DUBOW, Michael ; TITLE OF INVENTION: DNA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44 AHJD ; TITLE OF INVENTION: THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES ; FILE REFERENCE: 073406-0302 ; CURRENT APPLICATION NUMBER: US/09/727,892 ; CURRENT FILING DATE: 2000-12-01 ; NUMBER OF SEQ ID NOS: 159 ; SOFTWARE: PatentIn version 3.0 ; SEQ ID NO 1 ; LENGTH: 16668 ; TYPE: DNA ; ORGANISM: Staphylococcus aureus Bacteriophage 44 AHJD</p> <p>US-09-727-892-1</p> <p>Query Match 73.6%; Score 16.2; DB 12; Length 16668; Best Local Similarity 85.7%; Pred. No. 3.5e+02; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p> <p>QY 2 CGGGCAAAACCTCCGCTTAAAC 22 DB 6087 CGGTCAAACCGCGCCTAAC 6067 </p> <p>RESULT 28</p> <p>US-10-085-783A-15488</p> <p>; Sequence 15488, Application US/10085783A ; Publication No. US20040037841A1 ; GENERAL INFORMATION: ; APPLICANT: ChondroGene Inc. ; APPLICANT: Liaw, C.C. ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis ; FILE REFERENCE: 4231/2002 ; CURRENT APPLICATION NUMBER: US/10/085,783A ; CURRENT FILING DATE: 2002-02-28 ; PRIOR APPLICATION NUMBER: US 60/305,340 ; PRIOR FILING DATE: 2001-07-13 ; PRIOR APPLICATION NUMBER: US 60/275,017 ; PRIOR FILING DATE: 2001-03-12 ; PRIOR APPLICATION NUMBER: US 60/271,955 ; PRIOR FILING DATE: 2001-02-28 ; NUMBER OF SEQ ID NOS: 58994 ; SOFTWARE: PatentIn version 3.2 ; SEQ ID NO 15488 ; LENGTH: 158 ; TYPE: DNA ; ORGANISM: Human</p> <p>US-10-085-783A-15488</p>
---	---	--

Query Match 71.8%; Score 15.8; DB 13; Length 158;
Best Local Similarity 89.5%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTAA 21
| | | | | | | | | | | | | | | | | | | | | |
Db 127 GGGCAAAACCTCGACCTAA 145

RESULT 29

US-10-242-535A-15488
; Sequence 15488, Application US/10242535A
; Publication No. US20040013663A1

GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15488
; LENGTH: 158
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-15488

Query Match 71.8%; Score 15.8; DB 16; Length 158;
Best Local Similarity 89.5%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTAA 21
| | | | | | | | | | | | | | | | | | | | | |
Db 127 GGGCAAAACCTCGACCTAA 145

RESULT 30

US-10-106-698-1566/c
; Sequence 1566, Application US/10106698
; Publication No. US20030105690A1

GENERAL INFORMATION:

; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1566
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312)..(312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

; LOCATION: (314)..(314)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1566

Query Match 71.8%; Score 15.8; DB 15; Length 346;
Best Local Similarity 81.0%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCGCCTAA 21
| | | | | | | | | | | | | | | | | | | | | |
Db 319 CCGGGNANAACTCGGCCAAA 299

Search completed: June 20, 2004, 17:45:58
Job time : 136.434 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 705.989 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggcagggccgaattctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.in.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	26	100.0	401	6	AR149059	Sequence
C 2	26	100.0	401	6	AR340431	Sequence
C 3	26	100.0	18121	1	AE007182	Mycobacte
C 4	26	100.0	37164	1	MSGY126	Mycobacte
C 5	26	100.0	244800	1	EX842584	Mycobacte
C 6	26	100.0	278492	1	EX248347	Mycobacte
C 7	19.6	75.4	12461	1	AE007304	Sinorhizo
C 8	19.4	74.6	38793	1	MSGB27CS	L78817 Mycobacteri
C 9	19.4	74.6	348450	1	MLEPRTN4	AL583920 Mycobacte
C 10	18.6	71.5	505	6	AR226916	Sequence
C 11	18.6	71.5	1236	8	BT010447	BT010447 Arabidops
C 12	18.6	71.5	1766	3	BT011075	BT011075 Drosophil
C 13	18.6	71.5	1959	9	AB046101	AB046101 Macaca fa
C 14	18.6	71.5	2057	6	BD180518	BD180518 Highly th
C 15	18.6	71.5	2358	3	AF301149	AF301149 Drosophil
C 16	18.6	71.5	40970	3	LMFLUNK03	AL449123 Leishmani
C 17	18.6	71.5	42950	2	AC017471	AC017471 Drosophil
C 18	18.6	71.5	103904	8	ATAC021640	AC021640 Arabidops
C 19	18.6	71.5	110000	2	LMFLCHR31_13	Continuation (14 o
C 20	18.6	71.5	110000	2	LMFLCHR32_17	Continuation (18 o
C 21	18.6	71.5	158402	3	AC008343	AC008343 Drosophil
C 22	18.6	71.5	164840	2	AC011127	AC011127 Homo sapi
C 23	18.6	71.5	18553	9	AC022046	AC022046 Homo sapi
C 24	18.6	71.5	189412	2	AC023532	AC023532 Homo sapi
C 25	18.6	71.5	197837	2	AP001397	AP001397 Homo sapi
C 26	18.6	71.5	205152	9	AP002985	AP002985 Homo sapi
C 27	18.6	71.5	214701	2	AP001447	AP001447 Homo sapi
C 28	18.6	71.5	255652	3	AE003811	AE003811 Drosophil
C 29	18.6	71.5	301618	1	AE017177	AE017177 Porphyrom
C 30	18.6	71.5	344615	1	EX569695	EX569695 Synecoco
C 31	18.2	70.0	163025	14	AY372243	AY372243 Psittacidi
C 32	18.2	70.0	176539	9	AL354928	AL354928 Human DNA
C 33	18	69.2	18857	1	AE006947	AE006947 Mycobacte
C 34	18	69.2	22913	2	AC130497	AC130497 Rattus no
C 35	18	69.2	262082	2	AC120441	AC120441 Rattus no
C 36	18	69.2	315876	2	AC121279	AC121279 Rattus no
C 37	18	69.2	324050	1	EX248335	EX248335 Mycobacte
C 38	18	69.2	325841	2	AC094665	AC094665 Rattus no
C 39	18	69.2	327425	2	AC105666	AC105666 Rattus no
C 40	18	69.2	337436	2	AC098285	AC098285 Rattus no
C 41	18	69.2	342416	1	EX842573	EX842573 Mycobacte
C 42	18	69.2	343250	1	AP003594	AP003594 Nostoc sp
C 43	17.8	68.5	298700	1	AP005951	AP005951 Bradyrhiz
C 44	17.6	67.7	9973	1	AE000070	AE000070 Rhizobium
C 45	17.6	67.7	11159	1	AE012066	AE012066 Xanthomon
C 46	17.6	67.7	38141	2	AC018202	AC018202 Drosophil
C 47	17.6	67.7	82849	3	AC004277	AC004277 Drosophil
C 48	17.6	67.7	100589	8	OSJN00191	AL662991 Oryza sat
C 49	17.6	67.7	110000	2	LMFLCHR31_13	Continuation (14 o
C 50	17.6	67.7	110000	2	LMFLCHR31_14	Continuation (15 o
C 51	17.6	67.7	137332	8	AP002820	AP002820 Oryza sat
C 52	17.6	67.7	143710	8	AP002899	AP002899 Oryza sat
C 53	17.6	67.7	146840	8	AP003343	AP003343 Oryza sat
C 54	17.6	67.7	151203	8	CNS09S4W	EX255875 Oryza sat
C 55	17.6	67.7	153064	8	CNS08CAL	AL831808 Oryza sat
C 56	17.6	67.7	165512	3	AC099010	AC099010 Drosophil
C 57	17.6	67.7	175667	9	AC009336	AC009336 Homo sapi
C 58	17.6	67.7	245560	1	AP005330	AP005330 Vibrio vu
C 59	17.6	67.7	270766	3	AE003615	AE003615 Drosophil
C 60	17.6	67.7	298550	1	AP005961	AP005961 Bradyrhiz
C 61	17.6	67.7	300880	1	AE016917	AE016917 Chromobac
C 62	17.6	67.7	301235	1	AE016799	AE016799 Vibrio vu
C 63	17.6	67.7	301838	1	AE017209	AE017209 Geobacter
C 64	17.6	67.7	314250	1	EX294149	EX294149 Pirellula
C 65	17.6	67.7	320040	6	A79350	A79350 Sequence 1

C 66	17.6	67.7	320040	6	A93002	A93002 Sequence 1	C 139	17	65.4	146500	2	AC119500	Leishmani
C 67	17.6	67.7	348971	1	BX572594	BX572594 Rhodopseu	140	17	65.4	147844	2	AC068188	Homo sapi
C 68	17.4	66.9	1970	8	AY163775	AY163775 Arabidops	141	17	65.4	154541	8	AP003278	Oryza sat
C 69	17.4	66.9	1982	8	AY163773	AY163773 Arabidops	142	17	65.4	159843	5	EX000485	BX000485 Zebrafish
C 70	17.4	66.9	2004	8	AY163774	AY163774 Arabidops	143	17	65.4	163671	8	AP003330	Oryza sat
C 71	17.2	66.2	13230	1	AE014659	AE014659 Bifidobac	144	17	65.4	166757	8	AC098565	Oryza sat
C 72	17.2	66.2	13632	1	AE000851	AE000851 Methanoba	145	17	65.4	166757	8	AC098565	Oryza sat
C 73	17.2	66.2	161787	1	AC105679	AC105679 Rattus no	C 146	17	65.4	170523	9	AP002387	AP002387 Homo sapi
C 74	17.2	66.2	223820	2	AC094762	AC094762 Rattus no	147	17	65.4	171415	2	AP002771	AP002771 Homo sapi
C 75	17.2	66.2	225245	2	AC096910	AC096910 Rattus no	C 148	17	65.4	177577	3	AC022349	AC022349 Drosophil
C 76	17.2	66.2	229694	2	AC128965	AC128965 Rattus no	C 149	17	65.4	179016	3	AC018489	AC018489 Drosophil
C 77	17.2	66.2	274692	2	AC128068	AC128068 Rattus no	C 150	17	65.4	187795	2	AC025405	AC025405 Homo sapi
C 78	17.2	66.2	348014	1	BX640430	BX640430 Bordetell	151	17	65.4	188561	2	AC110739	AC110739 Mus muscu
C 79	17.2	66.2	349876	1	BX640442	BX640442 Bordetell	C 152	17	65.4	192001	2	AC068376	AC068376 Homo sapi
C 80	17.2	66.2	349980	6	AX492782	AX492782 Sequence	C 153	17	65.4	192372	2	AC110301	AC110301 Homo sapi
C 81	17.2	66.2	349980	6	AX553949	AX553949 Sequence	C 154	17	65.4	193314	2	AL603865	AL603865 Mus muscu
C 82	17	65.4	290	3	AY130267	AY130267 Microcton	C 155	17	65.4	213257	2	AC133102	AC133102 Mus muscu
C 83	17	65.4	295	3	AY130268	AY130268 Microcton	C 156	17	65.4	218800	2	AC094410	AC094410 Rattus no
C 84	17	65.4	304	3	AY130265	AY130265 Microcton	C 157	17	65.4	218912	2	AC128733	AC128733 Rattus no
C 85	17	65.4	315	3	AF428043	AF428043 Microcton	C 158	17	65.4	224371	2	AC132456	AC132456 Mus muscu
C 86	17	65.4	329	3	AF428045	AF428045 Microcton	C 159	17	65.4	292900	1	AC097163	AC097163 Rattus no
C 87	17	65.4	329	3	AY130263	AY130263 Microcton	C 160	17	65.4	295150	1	SCO919125	SCO919125 Streptomy
C 88	17	65.4	329	3	AY130264	AY130264 Microcton	C 161	17	65.4	299925	1	AP005042	AP005042 Streptomy
C 89	17	65.4	329	3	AY130266	AY130266 Microcton	C 162	17	65.4	300029	8	AE017057	AE017057 Oryza sat
C 90	17	65.4	334	3	AF428044	AF428044 Microcton	C 163	17	65.4	300029	8	AE017057	AE017057 Oryza sat
C 91	17	65.4	443	1	AF421322	AF421322 Unculture	C 164	17	65.4	300469	3	AE003503	AE003503 Drosophil
C 92	17	65.4	457	3	ACABETU	L31802 Acanthamoeb	C 165	17	65.4	300550	1	AP005030	AP005030 Streptomy
C 93	17	65.4	457	3	ACABETU	L31803 Acanthamoeb	C 166	17	65.4	302007	1	SCO939132	SCO939132 Streptomy
C 94	17	65.4	457	3	ACABETU	L31804 Acanthamoeb	C 167	17	65.4	303050	1	EX321860	EX321860 Nitrosomo
C 95	17	65.4	461	8	TTR555617	ATJ555617 Trameces	C 168	17	65.4	310029	1	AE016874	AE016874 Pseudomon
C 96	17	65.4	612	8	RHTBETU	L47275 Rhodotomula	C 169	17	65.4	310967	1	AE016874	AE016874 Pseudomon
C 97	17	65.4	706	11	PM12B2G	AL684314 Penicilli	C 170	17	65.4	344050	1	MLEBPTN1	MLEBPTN1 Mycobacte
C 98	17	65.4	709	11	PMGGHG	AL685021 Penicilli	C 171	17	65.4	347137	1	EX640448	EX640448 Bordetell
C 99	17	65.4	725	6	A86149	A86149 Sequence 80	C 172	17	65.4	349305	1	EX640433	EX640433 Bordetell
C 100	17	65.4	725	6	AE155642	AE155642 Sequence	C 173	17	65.4	349346	1	EX640433	EX640433 Bordetell
C 101	17	65.4	725	6	E66167	E66167 Genome DNA	C 174	17	65.4	349380	6	AX523952	AX523952 Sequence
C 102	17	65.4	879	6	BD251000	BD251000 Nucleic a	C 175	17	65.4	349980	6	AX523952	AX523952 Sequence
C 103	17	65.4	879	6	AR409440	AR409440 Sequence	C 176	16.8	64.6	17855	6	AX654077	AX654077 Sequence
C 104	17	65.4	1228	8	AY325811	AY325811 Alternari	C 177	16.8	64.6	2303	8	AK072823	AK072823 Oryza sat
C 105	17	65.4	1233	3	AF205379	AF205379 Leishmani	C 178	16.8	64.6	3141	6	E01473	E01473 DNA encod
C 106	17	65.4	1233	3	RHMNIFB	M15544 R.melliloti	C 179	16.8	64.6	5273	6	E01449	E01449 Genomic DNA
C 107	17	65.4	1942	3	AY069706	AY069706 Drosophil	C 180	16.8	64.6	5274	1	PSEBDO	PSEBDO Pseudomon
C 108	17	65.4	1951	8	AK067034	AK067034 Oryza sat	C 181	16.8	64.6	5274	6	B01338	B01338 Genomic DNA
C 109	17	65.4	2198	1	STMREDD	M29790 S.coelicolo	C 182	16.8	64.6	6716	1	PSETQDCLC	PSETQDCLC Putida to
C 110	17	65.4	2913	3	CEU132700	CEU132700 Caenorhab	C 183	16.8	64.6	11156	1	AF180147	AF180147 Pseudomon
C 111	17	65.4	3546	3	CEU132701	CEU132701 Caenorhab	C 184	16.8	64.6	15167	1	PFU18245	PFU18245 Pseudomon
C 112	17	65.4	5913	1	RER292535	AY292535 Rhodococc	C 185	16.8	64.6	77207	2	AC074067	AC074067 Homo sapi
C 113	17	65.4	6027	2	AC013143	AC013143 Drosophil	C 186	16.8	64.6	100800	8	AC105260	AC105260 Oryza sat
C 114	17	65.4	6971	1	AE004495	AE004495 Pseudomon	C 187	16.8	64.6	134247	8	AP004158	AP004158 Oryza sat
C 115	17	65.4	9979	1	AE004495	AE004495 Pseudomon	C 188	16.8	64.6	142854	8	AC136216	AC136216 Oryza sat
C 116	17	65.4	11610	1	AE004497	AE004497 Pseudomon	C 189	16.8	64.6	152813	2	AP005860	AP005860 Oryza sat
C 117	17	65.4	11632	1	AE014727	AE014727 Bifidobac	C 190	16.8	64.6	152813	2	AC141827	AC141827 Apis mell
C 118	17	65.4	12016	1	AE011702	AE011702 Xanthomon	C 191	16.8	64.6	187916	2	AP003458	AP003458 Oryza sat
C 119	17	65.4	12839	1	AE004617	AE004617 Pseudomon	C 192	16.8	64.6	192685	1	AE001863	AE001863 Deinococc
C 120	17	65.4	13187	1	AE005701	AE005701 Caulobact	C 193	16.8	64.6	198677	1	AE001863	AE001863 Deinococc
C 121	17	65.4	37566	1	MSGB971CS	L78821 Mycobacteri	C 194	16.8	64.6	203868	9	AP002381	AP002381 Homo sapi
C 122	17	65.4	42432	1	AF083072	AF083072 Cenarchae	C 195	16.8	64.6	261160	2	AC140208	AC140208 Mus muscu
C 123	17	65.4	42432	6	BD250982	BD250982 Nucleic a	C 196	16.8	64.6	262129	2	AC097899	AC097899 Rattus no
C 124	17	65.4	42432	6	AR409422	AR409422 Sequence	C 197	16.8	64.6	346837	1	AP002995	AP002995 Mesorhizo
C 125	17	65.4	70206	2	AC126804	Continuation (4 of	C 198	16.8	64.6	346837	1	AP002995	AP002995 Mesorhizo
C 126	17	65.4	87407	8	AP006369	AP006369 Lotus cor	C 199	16.8	64.6	349315	1	EX572593	EX572593 Rhodopseu
C 127	17	65.4	93483	6	AE695974	AE695974 Sequence	C 200	16.6	63.8	447	8	CPA345656	CPA345656 Cyanophor
C 128	17	65.4	103146	9	HS187N21	Z98036 Human DNA s	C 201	16.6	63.8	762	6	AR386170	AR386170 Sequence
C 129	17	65.4	110000	2	AC138524	AC138524 Homo sapi	C 202	16.6	63.8	889	3	PHTRNAGE	PHTRNAGE Phycophthor
C 130	17	65.4	110960	3	CEY39A1A	AL031633 Caenorhab	C 203	16.6	63.8	965	3	AF468039	AF468039 Oikopleur
C 131	17	65.4	112031	9	AL445487	AL445487 Human DNA	C 204	16.6	63.8	1053	1	AK073613	AK073613 Oryza sat
C 132	17	65.4	120330	8	AP004361	AP004361 Oryza sat	C 205	16.6	63.8	1053	1	AF063422	AF063422 Pseudomon
C 133	17	65.4	129757	8	ATP2K35	AL132956 Arabidops	C 206	16.6	63.8	1223	3	FSOCYHYC	FSOCYHYC Fusarium la
C 134	17	65.4	136197	9	AC105925	AC105925 Homo sapi	C 207	16.6	63.8	1305	6	AR389262	AR389262 Sequence
C 135	17	65.4	137560	6	AX754989	AX754989 Sequence	C 208	16.6	63.8	1650	3	AY118740	AY118740 Drosophil
C 136	17	65.4	140219	10	AL603714	AL603714 Mouse DNA	C 209	16.6	63.8	1827	8	CRU40465	CRU40465 Chlamydomon
C 137	17	65.4	141293	2	AP004337	AP004337 Oryza sat	C 210	16.6	63.8	2632	14	AF224760	AF224760 Cucurbit
C 138	17	65.4	146199	2	AP005107	AP005107 Oryza sat	C 211	16.6	63.8	2632	14	AF2566200	AF2566200 Cucurbit

AX598870	Sequence	285	16.6	63.8	200062	4	AC091623	Sus scrofa
AX647425	Sequence	286	16.6	63.8	200441	2	AC022130	Homo sapi
CRU42979	Chlamydomon	287	16.6	63.8	206486	2	AC074218	Mus muscu
AX346757	Sequence	288	16.6	63.8	215241	8	AF459639	Triticum
AX767474	Sequence	289	16.6	63.8	229887	2	AC073297	Mus muscu
AX822332	Sequence	290	16.6	63.8	232080	2	AC145374	Mus muscu
AX825972	Sequence	291	16.6	63.8	233000	1	AP000003	Pyrococu
UJ249601	Homo sapi	292	16.6	63.8	23356	3	AE003741	Drosophil
UJ1683	Brucella ab	293	16.6	63.8	243829	2	AC094446	Rattus no
AF263012	Streptomy	294	16.6	63.8	257621	2	AC094703	Rattus no
AE005683	Caulobact	295	16.6	63.8	260050	1	SM591782	Sinorhizo
AE005269	Escherich	296	16.6	63.8	280285	2	AC132158	Rattus no
AE005144	Halobacte	297	16.6	63.8	293050	1	SC0939116	Streptomy
AE009524	Brucella	298	16.6	63.8	297816	1	AP002553	Escherich
AE004795	Pseudomon	299	16.6	63.8	300000	1	SM591784	Sinorhizo
AE000189	Escherich	300	16.6	63.8	300000	1	SM591784	Sinorhizo
AE012109	Xanthomon					8	AE017076	Oryza sat
AE014416	Brucella							
AE005170	Halobacte							
AE014811	Bifidobac							
AE002035	Deinococc							
D90725	Escherichia							
AE007198	Sinorhizo							
AE012910	Chlorobiu							
AE002038	Deinococc							
AE005064	Halobacte							
BD250022	Insect p5							
Z79692	R.mellioti							
AX251534	Sequence							
AB047548	Pseudomon							
EX842628	Neutrosop							
AX695692	Sequence							
AC095019	Takifugu							
AC138753	Homo sapi							
AC141661	Apis mell							
AC019980	Drosophil							
AC020220	Drosophil							
AL713958	Mouse DNA							
Continuation (13 o								
AP003806	Oryza sat							
AC145181	Takifugu							
AC131374	Oryza sat							
AC091849	Homo sapi							
AC134232	Oryza sat							
AC079218	Mus muscu							
AC091749	Oryza sat							
AP003609	Oryza sat							
AC138625	Homo sapi							
AP005799	Oryza sat							
AX492788	Sequence							
AX553955	Sequence							
AC137788	Homo sapi							
AP003221	Oryza sat							
AC008200	Drosophil							
AP003232	Oryza sat							
BX546502	Danio rer							
AC008349	Drosophil							
BX295541	Homo sapi							
AC008199	Drosophil							
AC091933	Homo sapi							
AC005714	Drosophil							
AL627312	Mouse DNA							
AC091503	Rattus no							
AC018641	Human Chr							
AC132645	Rattus no							
AC112194	Homo sapi							
AC022505	Homo sapi							
AF016485	Halobacte							
AF016485	Halobacte							
AC026748	Homo sapi							
AL646058	Ralstonia							
AC084125	Homo sapi							
AB088420	Pseudomon							

AX598870	Sequence	285	16.6	63.8	200062	4	AC091623	Sus scrofa
AX647425	Sequence	286	16.6	63.8	200441	2	AC022130	Homo sapi
CRU42979	Chlamydomon	287	16.6	63.8	206486	2	AC074218	Mus muscu
AX346757	Sequence	288	16.6	63.8	215241	8	AF459639	Triticum
AX767474	Sequence	289	16.6	63.8	229887	2	AC073297	Mus muscu
AX822332	Sequence	290	16.6	63.8	232080	2	AC145374	Mus muscu
AX825972	Sequence	291	16.6	63.8	233000	1	AP000003	Pyrococu
UJ249601	Homo sapi	292	16.6	63.8	23356	3	AE003741	Drosophil
UJ1683	Brucella ab	293	16.6	63.8	243829	2	AC094446	Rattus no
AF263012	Streptomy	294	16.6	63.8	257621	2	AC094703	Rattus no
AE005683	Caulobact	295	16.6	63.8	260050	1	SM591782	Sinorhizo
AE005269	Escherich	296	16.6	63.8	280285	2	AC132158	Rattus no
AE005144	Halobacte	297	16.6	63.8	293050	1	SC0939116	Streptomy
AE009524	Brucella	298	16.6	63.8	297816	1	AP002553	Escherich
AE004795	Pseudomon	299	16.6	63.8	300000	1	SM591784	Sinorhizo
AE000189	Escherich	300	16.6	63.8	300000	1	SM591784	Sinorhizo
AE012109	Xanthomon					8	AE017076	Oryza sat
AE014416	Brucella							
AE005170	Halobacte							
AE014811	Bifidobac							
AE002035	Deinococc							
D90725	Escherichia							
AE007198	Sinorhizo							
AE012910	Chlorobiu							
AE002038	Deinococc							
AE005064	Halobacte							
BD250022	Insect p5							
Z79692	R.mellioti							
AX251534	Sequence							
AB047548	Pseudomon							
EX842628	Neutrosop							
AX695692	Sequence							
AC095019	Takifugu							
AC138753	Homo sapi							
AC141661	Apis mell							
AC019980	Drosophil							
AC020220	Drosophil							
AL713958	Mouse DNA							
Continuation (13 o								
AP003806	Oryza sat							
AC145181	Takifugu							
AC131374	Oryza sat							
AC091849	Homo sapi							
AC134232	Oryza sat							
AC079218	Mus muscu							
AC091749	Oryza sat							
AP003609	Oryza sat							
AC138625	Homo sapi							
AP005799	Oryza sat							
AX492788	Sequence							
AX553955	Sequence							
AC137788	Homo sapi							
AP003221	Oryza sat							
AC008200	Drosophil							
AP003232	Oryza sat							
BX546502	Danio rer							
AC008349	Drosophil							
BX295541	Homo sapi							
AC008199	Drosophil							
AC091933	Homo sapi							
AC005714	Drosophil							
AL627312	Mouse DNA							
AC091503	Rattus no							
AC018641	Human Chr							
AC132645	Rattus no							
AC112194	Homo sapi							
AC022505	Homo sapi							
AF016485	Halobacte							
AF016485	Halobacte							
AC026748	Homo sapi							
AL646058	Ralstonia							
AC084125	Homo sapi							
AB088420	Pseudomon							

AX598870	Sequence	285	16.6	63.8	200062	4	AC091623	Sus scrofa
AX647425	Sequence	286	16.6	63.8	200441	2	AC022130	Homo sapi
CRU42979	Chlamydomon	287	16.6	63.8	206486	2	AC074218	Mus muscu
AX346757	Sequence	288	16.6	63.8	215241	8	AF459639	Triticum
AX767474	Sequence	289	16.6	63.8	229887	2	AC073297	Mus muscu
AX822332	Sequence	290	16.6	63.8	232080	2	AC145374	Mus muscu
AX825972	Sequence	291	16.6	63.8	233000	1	AP000003	Pyrococu
UJ249601	Homo sapi	292	16.6	63.8	23356	3	AE003741	Drosophil
UJ1683	Brucella ab	293	16.6	63.8	243829	2	AC094446	Rattus no
AF263012	Streptomy	294	16.6	63.8	257621	2	AC094703	Rattus no
AE005683	Caulobact	295	16.6	63.8	260050	1	SM591782	Sinorhizo
AE005269	Escherich	296	16.6	63.8	280285	2	AC132158	Rattus no
AE005144	Halobacte	297	16.6	63.8	293050	1	SC0939116	Streptomy
AE009524	Brucella	298	16.6	63.8	297816	1	AP002553	Escherich
AE004795	Pseudomon	299	16.6	63.8	300000	1	SM591784	Sinorhizo
AE000189	Escherich	300	16.6	63.8	300000	1	SM591784	Sinorhizo
AE012109	Xanthomon					8	AE017076	Oryza sat
AE014416	Brucella							
AE005170	Halobacte							
AE014811	Bifidobac							
AE002035	Deinococc							
D90725	Escherichia							
AE007198	Sinorhizo							
AE012910	Chlorobiu							
AE002038	Deinococc							
AE005064	Halobacte							
BD250022	Insect p5							
Z79692	R.mellioti							
AX251534	Sequence							
AB047548	Pseudomon							
EX842628	Neutrosop							
AX695692	Sequence							
AC095019	Takifugu							
AC138753	Homo sapi							
AC141661	Apis mell							
AC019980	Drosophil							
AC020220	Drosophil							
AL713958	Mouse DNA							
Continuation (13 o								
AP003806	Oryza sat							
AC145181	Takifugu							
AC131374	Oryza sat							
AC091849	Homo sapi							
AC134232	Oryza sat							
AC079218	Mus muscu							
AC091749	Oryza sat							
AP003609	Oryza sat							
AC138625	Homo sapi							
AP005799	Oryza sat							
AX492788	Sequence							
AX553955	Sequence							
AC137788	Homo sapi							
AP003221	Oryza sat							
AC008200	Drosophil							
AP003232	Oryza sat							
BX546502	Danio rer							
AC008349	Drosophil							
BX295541	Homo sapi							
AC008199	Drosophil							
AC091933	Homo sapi							
AC005714	Drosophil							
AL62731								

ALIGNMENTS

RESULT 1
LOCUS AB149059 401 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 75 from patent US 6228371.
ACCESSION AB149059
VERSION AB149059.1 GI:15113650
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 401)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 75 08-MAY-2001;
FEATURES
Location/Qualifiers
1..401
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 100.0%; Score 26; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGCCCGAATCTACGAA 26
Db 148 ACAACGGCGAGCCCGAATCTACGAA 123

RESULT 2
LOCUS AR340431 401 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 75 from patent US 6572865.
ACCESSION AR340431
VERSION AR340431.1 GI:33731979
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 401)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides and methods for using same
JOURNAL Patent: US 6572865-A 75 03-JUN-2003;
FEATURES
Location/Qualifiers
1..401
/organism="unknown"
/mol_type="genomic DNA"

Query Match 100.0%; Score 26; DB 6; Length 401;

Best Local Similarity 100.0%; Pred. No. 0.45; Mismatches 0; Indels 0; Gaps 0;
 Matches 26; Conservative 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26
 Db 148 ACAACGGCGAGGCCGGAATCTACGAA 123

RESULT 3
 AE007182 18121 bp DNA linear BCT 27-APR-2001
 LOCUS Mycobacterium tuberculosis CDC1551, section 268 of 280 of the
 DEFINITION complete genome.
 ACCESSION AE007182 AE000516
 VERSION AE007182.1 GI:13883751
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis CDC1551
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 1 (bases 1 to 18121)
 Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
 Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
 Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
 Salzberg,S.L., Delcher,A., Uterback,T., Weidman,J., Khouri,H.,
 Gill,J., Mikula,A. and Bishai,W.
 Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains
 Unpublished
 2 (bases 1 to 18121)
 Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
 Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
 Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
 Salzberg,S.L., Delcher,A., Uterback,T., Weidman,J., Khouri,H.,
 Gill,J., Mikula,A. and Bishai,W.
 Direct Submission
 Submitted (25-APR-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..18121
 /organism="Mycobacterium tuberculosis CDC1551"
 /mol_type="genomic DNA"
 /strain="CDC1551"
 /db_xref="taxon:83331"
 /note="clinical strain"
 207..896
 /gene="MT3873"
 207..896
 /gene="MT3873"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48237.1"
 /db_xref="GI:13883752"
 /translation="MSAFDSGLRTGIVYVARPNWVANVTSMIDAAAGLHPRV
 ALMDVSGNPPGGSWNRLNLYADYAGSPVRIIGYANAYDFFNWRVPAGLR
 VIGAGYGNLPQVAHQYTDGSGYSPNLPGQAPFGRCNYSANGLTPOQFAAACG
 VTTTGGPLMALTDDEQTELTIKAREIWDQIRGPNAGWPQLGQNEQGLTPTVDIAIV
 IKNDVAAMLAIE"
 complement(910..1854)
 /gene="MT3874"
 complement(910..1854)
 /gene="MT3874"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK48238.1"
 /db_xref="GI:13883753"
 /translation="MPTDNDSWAITESVCATAGVAARAAETSDNPLNDPFARI
 FVDAAGDGWSMTNRTLLAGATDLDPLRAPIQQMIDFMAARTAFFEYFIATADAG

VRQVILLASGLDSAWRLPWPDDGTAVVYELDQPKVLEFKSATLRQHQAPASQLVNVP
 DLKQWPKALQAGDFSKPCWLAELGLVRYLPARAQDLLEFIDALSRLGSGWLASNV
 PGAGFLDERRRQRADNRNRAAAKLVETEISDVDDLWYAEQRTAVAEWLRERGWD
 VSTATPELLARYGRSIPHSGEDSIPPNFLFVSAQRTS"
 1884..2343
 /gene="MT3875"
 1884..2343
 /gene="MT3875"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48239.1"
 /db_xref="GI:13883754"
 /translation="MGSTPRTPOEFAHGHQALAGDLDEIVADYADDSEFVLT
 ARGKRGIRQLFVKLLDDIFNALMDLTKIFEGDILFLEWTANSASVRVDDGVTFVPR
 DGTIWAHTVRYTPHPT"
 complement(2388..2906)
 /gene="MT3876"
 complement(2388..2906)
 /gene="MT3876"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48240.1"
 /db_xref="GI:13883755"
 /translation="MTSPGTTNHNVSRIHRPOITVCRACKIAQPGESLIVSGAEOK
 VLDGLSLFQRPHPRIQVGLGRELADFGHVAQSSGVFPDGRDLSLVARVDGSE
 DGSIVGLVRFQSGDPCSKLLQRRHGLILSGRSAGSSAKCCNLTFSCAARNAPPGSI
 RPRQTTVPRSDR"
 2996..3127
 /gene="MT3876.1"
 2996..3127
 /gene="MT3876.1"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48241.1"
 /db_xref="GI:13883756"
 /translation="MNRTMGCADIGRWYGLGVADPVRARGRHPSPDVEDRALITWR"
 complement(3114..3689)
 /gene="MT3877"
 complement(3114..3689)
 /gene="MT3877"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48242.1"
 /db_xref="GI:13883757"
 /translation="MLSGIQNTLMDNDPLAHGYVADLLVALAVVLMRLARTRPE
 LARMLLCTGLVWELPVLGSAWNTPIEWATPLPTVPLVFLAHVYWDGPLLTM
 GWLLARALTGEAGALGLTVQVLWGQLTALVALSAILAGTWSYVDDLWPNVFWFR
 GHPTAAQMTLWLLPLCEALVRLALTAR"
 complement(3795..3977)
 /gene="MT3878"
 complement(3795..3977)
 /gene="MT3878"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48243.1"
 /db_xref="GI:13883758"
 /translation="MGSTPWPNCPCQCTLTPTVEVLEVALALPENPDRTAGAIQRL
 RAQLAGDRIALRGRGS"
 complement(3992..4183)
 /gene="MT3879"
 complement(3992..4183)
 /gene="MT3879"
 /note="identified by Glimmer2; putative"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48244.1"
/db_xref="GI:13883759"
/translation="MPAPKALSGVGFPRRIADLARPAAETVGRWLRPFAAEAVRS
PFRKVRISRHTIDRWIRN"
complement(4318..4644)
/gene="MT3880"
complement(4318..4644)
/gene="MT3880"
complement(4318..4644)
/gene="MT3880"
complement(4318..4644)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK48245.1"
/db_xref="GI:13883760"
/translation="MPAPKALSGVGFPRRIADLARPAAETVGRWLRPFAAEAVRS
VFTVMLRVDPPVMPDAAVGFAYAVTVAIVVIEQFALSTVSLAETAVAVSGG
RLVAPG"
complement(4779..4851)
/gene="trNA-Arg-4"
complement(4779..4851)
/gene="trNA-Arg-4"
complement(4779..4851)
/product="trNA-Arg"
complement(4882..4970)
/gene="trNA-Ser-3"
complement(4882..4970)
/gene="trNA-Ser-3"
complement(4882..4970)
/product="trNA-Ser"
5048..6109
/gene="MT3881"
5048..6109
/gene="MT3881"
/notes="similar to GP:5712092; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="histidinol-phosphate aminotransferase"
/protein_id="AAK48246.1"
/db_xref="GI:13883761"
/translation="MTARLRPELAGLPVYVPGTKVGAIKLASNETVFGPLSPVRAAI
DRATVNRYPONGVQLKAALRHLGPDFAPEHVAVGSGVSLCQQLVQVTSVSGDE
WVGMFSFLYPPQVRVAGAIPIQVPLTDHFDLVAMLATVTDTRLIFVCPNPNPTS
TVVGPALARFVEAPAHILLIADAEAYEYIRDGMRPDSGLVRAHNNVVLRTSKA
YGLAGURIGTGAIGHPDVTALDKVYVFFTVSSIQQAASLADDELLARTDTVVAE
RAVSAELRAAGFTLPSPQANFVWLPLGSRTOQDFVEQAADARIVVRPYGTGVRVTVA
APEENDAFRLRFARWRSDQ"
complement(6155..6739)
/gene="MT3882"
complement(6155..6739)
/gene="MT3882"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK48247.1"
/db_xref="GI:13883762"
/translation="MPPERPDPSPPTDELACAEALQVQLQVLTHTIGROKAKQTP
CPGYDKKTEHLNLSIMVIGNVGAEFSIRADIDSVERLVSAGSAALDAWRHGLE
GDVSLGFGMSKAVSVSVFSEFLVHWADYAVAGSELKAADSLAEYVLELARKULKP
EERSVAGFNEPVDVDPDGGALERLIATGTGNPAR"
6763..7587
/gene="MT3883"

```

Query Match 100.0%; Score 26; DB 1; Length 18121;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26

Db 14841 ACAACGGCGAGGCCGGAATCTACGAA 14866

RESULT 4
 MSGY126
 LOCUS
 DEFINITION Mycobacterium tuberculosis sequence from clone Y126.
 ACCESSION AD000012.1 GI:1717733
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

MSGY126 37164 bp DNA linear BCT 10-DEC-1996
 Mycobacterium tuberculosis sequence from clone Y126.

AD000012.1 GI:1717733
 Mycobacterium tuberculosis

Myobacterium tuberculosis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

1 (bases 1 to 37164)

Du, L.

Direct Submission

Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100

Beaver Street, Waltham, MA, USA, 02154 du@cric.com

CSDB:S:1004704

Location/Qualifiers

1..37164

/organism="Mycobacterium tuberculosis"

/mol_type="genomic DNA"

/db_xref="taxon:1773"

/clone="Y126.nt"

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 37164;

Best Local Similarity 100.0%; Pred. No. 0.43;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26

Db 13576 ACAACGGCGAGGCCGGAATCTACGAA 13601

RESULT 5

MSGY126

LOCUS

DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 13/13.

ACCESSION BX842584 AL021426 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864

VERSION Z94121 Z97188

KEYWORDS BX842584.1 GI:38490370

SOURCE complete genome.

ORGANISM Mycobacterium tuberculosis H37Rv

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

REFERENCE 1

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,

Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,

Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,

Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,

Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,

Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,

Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,

Squires, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and

Barrell, B.G.

Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence

Nature 393 (6685), 537-544 (1998)

JOURNAL 98295987

MEDLINE 9634230

REFERENCE 2

Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.

Re-annotation of the genome sequence of Mycobacterium tuberculosis

H37Rv

Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)

JOURNAL 22255591

MEDLINE 12368430

REFERENCE 3

(bases 1 to 244800)

Parkhill, J.

TITLE	Direct Submission	gene	product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk	CDS	/protein_id="CAA18045.1"
COMMENT	On or before Nov 21, 2003 this sequence version replaced gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648, gi:3261687, gi:3261736, gi:3261805.		/db_xref="GI:2960147"
Notes:	Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.		/translation="MGRKAVLVHSAFSGIGAVLYFYFVLPWPELMGDTGHSIGTGLRIALGALVGLAALPVFTLLTRKPELGTQALSMRIWISIMHVLGALITVGAISEWLSLDAAGOMLVFGYIGAAAIAVLGFFGYLSFVAELPPPPKPKPKPKORLRKKTAKGDEAPEAAEAENTELAAQDEEVAEPESIESPGGPESTREAPAETAATPEPRGLNRRPTGKTSRRRTSRGQVQAKYDE"
FEATURES	Location/Qualifiers		2735..2977
Source	1..244800		/gene="cut5a"
	/organism="Mycobacterium tuberculosis H37Rv"		/locus_tag="RV3724A"
	/mol_type="genomic DNA"		2735..2977
	/strain="H37Rv"		/gene="cut5a"
gene	/db_xref="taxon:83332"		/locus_tag="RV3724A"
	complement(89..1396)		/EC_number="3.1.1.-"
CDS	/locus_tag="RV3722c"		/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."
	complement(89..1396)		/note="RV3724A, (MTV025.072), len: 80 aa. Probable cut5a, truncated cutinase precursor (EC 3.1.1.-), similar to N-terminal end of others e.g. Q9KX87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa overlap); Q9XB09 RVD2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200, E(): 1.5e-06, (61.4% identity in 57 aa overlap); and Q00298 CUT1 BOTCI CUTA CUTINASE PRECURSOR from Botrytis cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores: opt: 108, E(): 2.2 (40.4% identity in 52 aa overlap). Also highly similar to others from Mycobacterium tuberculosis e.g. Q06318 CUT3_MYCTU RV3451 MT3557 MTCY13E12.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E(): 1.2e-05, (58.0% identity in 50 aa overlap); Q50664 CUT2_MYCTU RV2301 MT2358 MTCY339.08c PROBABLE CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E(): 0.0015, (59.2% identity in 49 aa overlap); Q06793 RV1758 MTCY28.24 295890 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap); Q06319 RV3452 MTV13E12.05; and U00015.11 from Mycobacterium leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724B CUT5B, frameshifting could occur near position 4169668. Sequence has been checked but no errors found."
	/codon_start=1		/codon_start=1
	/evidence="experimental"		/transl_table=11
	/transl_table=11		/product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A"
	/product="CONSERVED HYPOTHETICAL PROTEIN"		/protein_id="CAA18044.2"
	/protein_id="CAA18044.2"		/db_xref="GI:38490371"
	/db_xref="GI:38490371"		/translation="MSFDSLSPQELALHARHODVYALQGMKLADLTRKPSAQLDLNQLSLPDDYRPEGDTDRNYGGQGLPGLRAIFAEELGIAVFNLIAGNSLELMHDIVAFMLYGVSPRPWIQEOGKIFLCPVPGYDRHFAITETMGEMIPIMLQDGPVDLIELVADPAIKGMWTPVFGNPSGVYSWETVRRLVQRTAAPDFRLFWIDNAVAVHTLIDFPDQVDVLGAAGKNRPVYFASKTITFAGGCVSFGGSLGNIAWYLYAGKKSIGDKPKNOLHLEFFGDAGVRLMLRHQOILLAPKELVAEVLQRLSEKIASWTEKGGYFISLVDLPQGTARTVALADKVGIAVIEAGASPYRKDPDDKNRIAPSPFSPDLRNVADGLATCALLAATETLLNGLASSAPNVK"
gene	1613..1698		/gene="cut5b"
trna	/gene="trna-Ser (GGA)"		/locus_tag="RV3724B"
	1613..1698		2874..3437
	/gene="trna-Ser (GGA)"		/locus_tag="RV3724B"
	/product="trna-Ser"		/gene="cut5b"
	/notes="codon recognized: UCC"		/locus_tag="RV3724B"
gene	/anticodon=(pos:1647)..1649,aa:Ser)		/EC_number="3.1.1.-"
CDS	1804..2568		/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)."
	/locus_tag="RV3723"		/note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end of others e.g. Q9XB09 RVD2-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 335, E(): 3.4e-12, (53.25% identity in 92 aa overlap); Q9KX87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. Q06793 RV1758 MTCY28.24 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.5e-29, (57.25% identity in 166 aa overlap); Q06319 RV3452 MTCY13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN (226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%


```
identity in 165 aa overlap);
O06318|CUT3 MYCTU|RV3451|MT3557|MTCY13E12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 307, E():
1.9e-10, (40.7% identity in 167 aa overlap);
Q10837|CUT1 MYCTU|RV1984C|MT2037|MTCY39.35 PROBABLE
CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 261, E():
6.7e-08, (50.9% identity in 169 aa overlap) etc; and
U00015.11 from Mycobacterium lepra. 5'-end of gene is
RV3724A|CUT5A; frameshifting may occur near position
4169668. Tbpase score is 0.918."
/codon_start=1
/transl_table=11
/product="PROBABLE CUTINASE [SECOND PART] CUT5B"
/protein_id="CAE55633.1"
/db_xref="GI:38490373"
/translation="MAPGSHLVLAASEDCSSCHVQVGAQSLGVYAVNPASNDFAF
SDPKTVIGDIGHIQSMAMGSPQTRVLGVSQGAAGVYTSVAVPPAVPQVA
VPAPVAPEVANVAAVTLFGAPSAQFLGQYGAIPPIAIGFLYQPKTLQLCADGDSICGD
GNSFVAHGLYAVNGVQGANFAASRL"

Query Match      100.0%; Score 26; DB 1; Length 244800;
Best Local Similarity 100.0%; Pred. NO. 0.43;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGAA 26
    |||||
Db 60195 ACAACGGCGAGCGCCGAATCTACGAA 60220

RESULT 6
BX248347
LOCUS
DEFINITION
    BX248347 278492 bp DNA linear BCT 11-JUN-2003
    Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
    14/14.
ACCESSION
    BX248347 BX248333
VERSION
    BX248347.1 GI:31620456
KEYWORDS
    complete genome.
SOURCE
    Mycobacterium bovis subsp. bovis AF2122/97
    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
    Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
    tuberculosis complex.
REFERENCE
    1
    Garnier,T., Eigmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
    Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
    Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
    Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
    Hewinson,G.
    The complete genome sequence of Mycobacterium bovis
    Online Publication
    PNAS 10.1073/pnas.1130426100 ( Microbiology )
    2 (bases 1 to 278492)
    Garnier,T.
    Direct Submission
    Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
    Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
    15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
    Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
    Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
    Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
    Hinxton,Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
    Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
    Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
    rue du Docteur Roux, 75724 Paris Cedex 15, France
    Location/Qualifiers
        1. 278492
            /organism="Mycobacterium bovis subsp. bovis AF2122/97"
            /mol_type="genomic DNA"
            /strain="AF2122/97"
            /db_xref="taxon:233413"
            complement(240..704)
            /locus_tag="Mb3713c"
            complement(240..704)

gene
CDS
```

```
/locus_tag="Mb3713c"
/notes="Mb3713c", len: 154 aa. Equivalent to RV3688c,
len: 154 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 154 aa overlap). Hypothetical protein,
similar to other bacterial hypothetical proteins e.g.
QX9334|SCH66.02c from Streptomyces coelicolor (154 aa),
FASTA scores: opt: 425, E(): 3.4e-19, (46.1% identity in
154 aa overlap); QWZP4|TW0690 from Thermotoga maritima
(149 aa), FASTA scores: opt: 326, E(): 3.4e-13, (40.4%
identity in 151 aa overlap); Q9PHU3|CU0573 from
Campylobacter jejuni (147 aa), FASTA scores: opt:290,
E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc.
Also some similarity to upstream
O69654|RV3686c|MTV025.034c CONSERVED HYPOTHETICAL PROTEIN
from Mycobacterium tuberculosis."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95899.1"
/db_xref="GI:31620457"
/translation="MAELSQLRSLDTQAMKTQKLTATIRMLLAIOTEEYSGKOA
REUSDDEVIKLARESKRGEAAEIVTONGRELAATEHAERIIDYLPPTTEGEL
ADVADTAIAEVAEELGHRPSKQMGVLAATYIAAGKADGARLSAAVKERL"
704..2059
/locus_tag="Mb37114"
/locus_tag="Mb37114"
/notes="Mb37114", len: 451 aa. Equivalent to RV3689, len:
451 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.8% identity in 451 aa overlap). Probable conserved
transmembrane protein, with Proline rich N-terminus,
similar to Q9KYW6|SCE33.17 PUTATIVE INTEGRAL MEMBRANE
PROTEIN from Streptomyces coelicolor (462 aa) FASTA
scores: opt: 730, E(): 2.7e-21, (38.1% identity in 412 aa
overlap)."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAD95900.1"
/db_xref="GI:31620458"
/translation="MHKRYAPQRPKPTQETVIEKTRDROGDHERRQLLRPVSMPLP
PGVPBPVPAPVAGYAPYAPYPPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPG
CGPPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPGYPG
VVTQIISLVALGPGMTAFQDITVGPDELDTGAVGVGWSAFGASLLVSLVGLSVA
VGLTVGRVFGSPITVGEAWKVRGLLALFGLALLEAGVAVVGLAVVLSGVA
AAANEAAALGLFLLVGVSLAYLVLLFAPLVILKPLVEALITRSFALVRHG
FWRVILGRLLTVLVGVGVAIAAFVIMGEIVTAVTASDGSVTMRVLGATLSAIGVT
IGQIVTAPFSAGVWVLLYTORRIRAEAFDLVLTQLEAGPAGGAPVVESTDLWLTRP
F"
2086..2739
/locus_tag="Mb3715"
2086..2739
/locus_tag="Mb3715"
/notes="Mb3715", len: 217 aa. Equivalent to RV3690, len:
217 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 217 aa overlap). Probable conserved
membrane protein, similar to Q9KYW5|SCE33.18 PUTATIVE
INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor
(231 aa), FASTA scores: opt: 419, E(): 1.5e-19, (36.0%
identity in 211 aa overlap). Equivalent to AAK49159 from
Mycobacterium tuberculosis strain CDC1551 (233 aa) but
shorter 16 aa."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD95901.1"
/db_xref="GI:31620459"
/translation="MPSIDIDREAAHQARQLDKPIYPKDSLTKELFTWIDEQYLR
LEKGSIPGWFITVLLILMIAVTAARVTRMTRNGDQQLFDAGQUTAAQHR
STAESYAAEGNWAARIRHLOAVARELEETGMLNPAAGRTANELASDAGEVLPHAGE
LTQATATNDVTYGERPQTQAYQMIALDDHLRSRSPAVVSQVQHPAFVDSQAQR"
2865..3866
/locus_tag="Mb3716"
/locus_tag="Mb3716"
```

CDS

2865...3866
 /locus tag="Mb3716"
 /note="Mb3716", len: 333 aa. Equivalent to Rv3691, len: 333 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 333 aa overlap). Conserved hypothetical protein, similar to Q9KYM4|SC33.19 PUTATIVE SECRETED PROTEIN from Streptomyces coelicolor (387 aa)
 PASTA scores: opt: 481, E(): 6e-23, (36.6% identity in 358 aa overlap). Equivalent to AAK48160 from Mycobacterium tuberculosis strain CDC1551 (381 aa) but shorter 48 aa.
 /codon_start=1
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD95902.1"
 /db_xref="GI:31620460"
 /translation="MAPASTSSSTGGHALATLLGNHVEVVVADSIADVEAARPDILL
 LVAOTYLDNALDRKAPGDLIVAPTSRTTALTPOLRIAAASPENQPCNTLL
 EANPAGVOMGSPSTYQATGDLVLTSCYGGALYRFAEGRTITVGGSHPTNGGLIP
 AGNALANLGNRPRLVITAPDHIEGMSPSLSDLIPENVHTWQLWLVLLVA
 LMKRRIGPLVAELPVIIRASSETVEGRGLYRSRRARDADATLQIRLRLPG
 VGAGAPAPVVTITIAQSKADPPFPVYHLFGPAPATDNDLLQARALDDIERQVHS"
 3863...4939
 /gene="moxr2"
 /locus tag="Mb3717"
 /note="Mb3717", len: 358 aa. Equivalent to Rv3692, len: 358 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 358 aa overlap). Probable moxr2, methanol dehydrogenase regulatory protein, highly similar (generally longer at N-terminus) to Q9KYM3|SC33.20 PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (329 aa), PASTA scores: opt: 1523, E(): 4.2e-74, (70.9% identity in 330 aa overlap); Q92538|SC9B2.21C PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (332 aa) PASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in 313 aa overlap); Q9UZ67|MOXR-3|PAB0848 METHANOL DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi (314 aa), PASTA scores: opt: 989, E(): 1.1e-45, (50.65% identity in 302 aa overlap); Q9AANI|CC0566 MOXR PROTEIN from Caulobacter crescentus (323 aa), PASTA scores: opt: 988, E(): 1.3e-45, (52.3% identity in 306 aa overlap); etc. Also similar to O53170|MTV007.26|MOXR|Rv1479 from Mycobacterium tuberculosis (377 aa); and O07392|AF002133.6|MOXR from Mycobacterium avium (309 aa). Also high similarity with several hypothetical bacterial proteins."
 /codon_start=1
 /transl_table=11
 /product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL REGULATORY PROTEIN MOXR2"
 /protein_id="CAD95903.1"
 /db_xref="GI:31620461"
 /translation="MTOSASNPQAPTPTGALPYPPOAGAPTAAPSGPHPHRAE
 AESARDALLAEVAKAVGQGVISGLVIALLCRHVLFGLGCVKATLIVRAMSA
 ALQLEPRVQFTDPLMGDVTGSDVATAEFVFPFVFNLLIADINLTPKTKO
 AALLEAMEERQGEKPLPNPFIIVAAQNPPIEYEQYQLPEQLDRFLKLVNLL
 PARDSEIADRHAGHFDRLDSAINPAGPAELAGREAVRHVLANVELGVIVDVI
 GATESPALGLVSPRGATALLGTARSNAWLSGRDVTPTDDVKANARPTLRHVMRLP
 EASLEGATPDGVDGLIASVPVPR"
 5073...6395
 /locus tag="Mb3718"
 /note="Mb3718", len: 440 aa. Equivalent to Rv3693, len: 440 aa (alternative start at 41910), from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 440 aa overlap). Possible conserved membrane protein, similar to Q9KYM2|SC33.21 PUTATIVE LIPOPROTEIN from Streptomyces coelicolor (436 aa), PASTA scores: opt: 875, E(): 3.3e-46, (56.25% identity in 448 aa overlap); Q9AANO|CC0567 HYPOTHETICAL PROTEIN from Caulobacter crescentus (437 aa),

gene

CDS

AE007304 12461 bp DNA linear BCT 15-AUG-2001
 Sinorhizobium meliloti plasmid pSymb section 110 of 121 of the complete plasmid sequence.
 ACCESSION AE007304 AE006469
 VERSION AE007304.1 GI:14524346
 KEYWORDS
 SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
 ORGANISM Sinorhizobium meliloti
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 REFERENCE 1 (bases 1 to 12461)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymb megaplasmid
 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
 JOURNAL 21396509
 MEDLINE PUBMED
 REFERENCE 2 (bases 1 to 12461)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Direct Submission
 Submitted (29-MAR-2001) Biological Sciences, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA
 LOCATION/Qualifiers
 FEATURES
 source
 1. 12461
 /organism="Sinorhizobium meliloti"
 /mol_type="genomic DNA"
 /strain="1021"
 /db_xref="taxon:382"
 /plasmid="pSymb"
 /complement(134).520
 /gene="Sma2191"
 /complement(134).520
 /gene="Sma2191"
 /function="Miscellaneous; Hypothetical/Global homology"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK65841.1"
 /db_xref="GI:14524347"
 /translation="MRVGRVEILGOERRRRWDGAKLDIVMSIGLDGATVTEVAHRH
 DVSEQQIYANRHEHKKGLLPSTADTVFLPVDVAMHGPLVREDVARYSAMIELRN

gene

CDS

Query Match 100.0%; Score 26; DB 1; Length 278492;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 DB 96222 ACAACGGCGAGCCCGCAATCTACGAA 96247
 |||||
 RESULT 7
 AE007304/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1. 12461
 /organism="Sinorhizobium meliloti"
 /mol_type="genomic DNA"
 /strain="1021"
 /db_xref="taxon:382"
 /plasmid="pSymb"
 /complement(134).520
 /gene="Sma2191"
 /complement(134).520
 /gene="Sma2191"
 /function="Miscellaneous; Hypothetical/Global homology"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK65841.1"
 /db_xref="GI:14524347"
 /translation="MRVGRVEILGOERRRRWDGAKLDIVMSIGLDGATVTEVAHRH
 DVSEQQIYANRHEHKKGLLPSTADTVFLPVDVAMHGPLVREDVARYSAMIELRN

```

DGRSLRFDGVDPAVLTIRAVEAA"
complement (575. .1888)
/gene="Sma2193"
CDS
complement (575. .1888)
/gene="Sma2193"
/function="Small molecule metabolism"
/notes="glimmer prediction; similar to agaE [Agrobacterium tumefaciens], gb|AA88469. Weak similarity to pyridine nucleotide-disulphide oxidoreductase family"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="AAK65842.1"
/db_xref="GI:14524348"
/translation="MTRTGIGPSQKWDLIIVGGIGVGSTAFYAARLGLRLVIVERD
TPGSGQGNLGVROQGRDFELPLAIALRLMNDLEDLGRKVGWLRGNNVLAAN
DRTAQAQWQAKDFGLDTLLTSAQTRKPLPLSDKASLCGAMFTASDGRAPGR
TTPAFAQWQAKDFGLDTLLTSAQTRKPLPLSDKASLCGAMFTASDGRAPGR
LACVGNLQERIRAVARTLPARGTLITSCISLPTGLRQDEKGFIEISVAGGYDV
RFDUSWLRHYRTRRANPDARVNTFGILRPMSSRIASPIADIAPSEGGARPAHYR
VROQDELQFVQLAEALAEAVAGVIDLDPVPMGVHYERPGLLLATGFGSHGF
GLGPMAGKIMAGLAAGHSASVDISGLSPRFSRLSVTARLASPAG"
complement (1889. .2653)
/gene="Sma2195"
CDS
complement (1889. .2653)
/gene="Sma2195"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to amino acid and opine transporters, e.g., Agrobacterium tumefaciens P35117. ABC transporter family"
/codon_start=1
/transl_table=11
/product="probable ABC transporter, ATP-binding protein"
/protein_id="AAK65843.1"
/db_xref="GI:14524349"
/translation="MSHPMIVDRVSKSYGSMVTMLNDINLEIATGTSVILIGPSGSGK
STLLRCVNHLEIRERSITVDGELIGYVRVGNLAHELPEKLVARQAGIGWFOFNL
FGHKTALVEIAPVHVRGLARDAEQATALLERVLADKQOSYPMLSGQOQORVA
IARLAKPKVILPDEPTSLADPELAGEVLDWRSIAAEGLTWVWVTHEMSFAREVCN
RIAPMAGEIVECAPSEIFDKPAQRTRSFLSKVH"
complement (2655. .3569)
/gene="Sma2197"
CDS
complement (2655. .3569)
/gene="Sma2197"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to permease AtrB from Streptomyces coelicolor, AAF25332; binding-protein-dependent transport system"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65844.1"
/db_xref="GI:14524350"
/translation="MSNDLTAAPTASDPRLTYVPASTWTKRFAATLLILLVLYCAT
MIASNPFGVDVADVLDLSRLIWLGLSLTVMTVTVMVIGVLTGTFPAMAMADNVVI
STVANAYWFRFTPTVLVQLIPWNLGALFPQLSVGVPFTSLWVPTTLSPVTPAA
VLGLNGEAYMSIEIIRSGMSVDPGQQAQSLGTMGTNGTWRILLPQAMPYIIPPT
GNOTIGMLKTSILSVTSIADLYSAQTIYSRNFQIPLLIIVACIWLAAATLISAVQ
VRTERFARSQRANITQPERLFRQAR"
complement (3633. .4523)
/gene="Sma2199"
CDS
complement (3633. .4523)
/gene="Sma2199"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; putative bacterial extracellular solute-binding protein, similar to Streptomyces coelicolor AtrA, AAF25531"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, periplasmic solute-binding protein"
/protein_id="AAK65845.1"
/db_xref="GI:14524351"
/translation="MSAVKANRYQTAVATAATWAVAGVVLTTGAVHAQTAADILPOLYR
DAGVILVTDAKYPPFQSVNDAGEIVGFVDLWNAIADRLNVRKVDVTSVFSFSLIPGV
QAGRWDIAMEGITDNAERQKVSFVDYGYTTSSAYVLEQSGSEIKDHLGICGKSAQ
SGTEWGMIAKEIGDCAVAAKQKPSVSEFGTSEATLLSLYSGRSDFLTSALAGEFI
QKVAHPVKVIVPMOILPRMPSGIAFRKDETDLDGALLALKLKEIRANGDYEKIVAKTV
SPMAVEHEPGINLATVPQTK"
complement (4635. .5015)
/gene="Sma2201"
CDS
complement (4635. .5015)
/gene="Sma2201"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction; some similarity to putative translation initiation inhibitors. similar to SMA2289 of S. melliloti pSymba."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65846.1"
/db_xref="GI:14524352"
/translation="WKXRIFAGVSEEMAGYAKAVIVGTTVVSGTTGRDPTDGNFPD
NAAQARNALASIDKALROAGGSANAVSVVVDQACADVTAVLGEVFRD.RPTS
TMLICEIPAPAKVEIEITASLEC"
complement (5106. .5900)
/gene="Sma2203"
CDS
complement (5106. .5900)
/gene="Sma2203"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to transport protein PotC [Actinobacillus actinomycetemcomitans]: gb|AAC27497"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65847.1"
/db_xref="GI:14524353"
/translation="MTRWDISSIALRLVTFSSMAFLIPIPIATVILVAVNPREFVLPPN
GLTDWFPRAAASSXTFLRGMGVSLVGLGAALIANALPAALAIARNTFRFGKAINL
LMTPLPTTLTILSALYIFVRVYSGGLVPLVGLVHAHMPYAVRILTFASLNFPD
SVEEAARNVAGRIITFSLVPTRTGLISLTLFCILSWNDPPISVFLAPPBWTPFL
PVELSYIKFOYDAVAALASSLILLSAVAMVWIDRLAGLRRLRS"
complement (5897. .6766)
/gene="Sma2205"
CDS
complement (5897. .6766)
/gene="Sma2205"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to PotB permease for spermidine and putrescine [E. coli]: P23860"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65848.1"
/db_xref="GI:14524354"
/translation="MPENTRILPLLSPALALLLFVAPMVWVFRFSGNELAEGAYI
VEAMTLENVTRFLTEPVLVNTLWFSVRIALATAISVICAPVAVYAKTSGLOQNV
LMALTMALILGLVTVYGVTVIPEGGGLMSLMALRVDPQVRYWMDIKGVVILLV
YIGTPIYVLSLDSIERINPFLVEARNVGNRWTFWKIVPELSTGLYAGLVVFS
LNFAFVFLMIGDSNTQMGILYITREALNNDLPFASALSVIMVSNALLILGMSAL
AGRLLSRLAKR"
complement (6759. .7844)
/gene="Sma2207"
CDS
complement (6759. .7844)
/gene="Sma2207"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to putative ABC-transporter AtrA required for attachment to host cells and for virulence [Agrobacterium tumefaciens]: AAB67294"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, ATP-binding protein"
/protein_id="AAK65849.1"
/db_xref="GI:14524355"
/translation="MSNDSIQVEGITKRYGAMTAVDNVSFVQGEFVSLIGPSCGCK
TTTLRIAGFVDPGGLLRVRGRDVTLLPPEKRDILGFVFQNYALWPHMTVAENVAFGL
```



```

Similar to ML0117, ML0596 and ML1708"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC31223.1"
/db_xref="GI:13092923"
/db_xref="GOA:O32975"
/db_xref="SWISS-PROT:O32975"
/translation="WVDVDDIQAARADPFILQETVNGKPLIFDNAAATQKPOVIDR
LSFYAHNSNIHRAHEAARATDAIEARETARFRGAAGAQAIIIFVGTTEAINL
VAYAMGRLQPDGVIIITLHSHANIVPWLSSQTGAALKVAPVDDAGNLLMSBEF
DLGPRTKLVATQVNSALGTVTOGEKI VELGHRYGARVLIDGAQSI PHLPINSELG
ADFFVFSGHKI VPTGTCIGLVGCEVLTMPFPWQGGNNI VDVTLERSLYOCPNSELG
AGTNADAGLEALGVYVERGVORIASHEOALDYATPRLADIPGVRLVGTATEKA
SVLSFVLAGEHLEVGKALNAEGIAVRAGHCAQPVLRRLGLEATVPSPAFNYTYES
IDVFNVRRIRAEQTNIG"
/misc_feature
653..1738
/genes="ML0842"
/notes="Pfam match to entry PF00266 aminotran 5,
Aminotransferases class-V, score 23.70, E-value 1.7e-08"
1908..2395
/genes="lppp"
/pseudo
/notes="synonym: ML0843"
1908..2395
/genes="lppp"
/notes="Possible pseudogene of M. tuberculosis orthologue
lppp (Best blastx score 294)"
/pseudo
/codon_start=1
/transl_table=11
/product="lipoprotein (pseudogene)"
2549..2555
/notes="possible RBS"
2559..4112
/genes="nark"
/notes="Synonym: ML0844"
2559..4112
/genes="nark"
/notes="Similar to Mycobacterium tuberculosis putative
nitrite extrusion protein narK1 or Rv2329c or MTCY3612.05
TR:P71683 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,
69.3% id in 488 aa and to Escherichia coli nitrite
extrusion protein 2 narU SW:NRU.ECOLI (P37738; P77696)
(462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There
is a frameshift near the C-terminus relative to the M.
tuberculosis homologue. Previously sequenced as TR:O32974
(EMBL:Z98741). Contains hydrophobic, probable
membrane-spanning regions."
/codon_start=1
/transl_table=11
/product="putative nitrite extrusion protein"
/protein_id="CAC31225.1"
/db_xref="GI:13092924"
/db_xref="SPTREMBL:O32974"
/translation="MDQLVLLQAEESLYGPRRSKTVPIDRHSLKSSFFRLRRAC
RISHDPEDQAEAGNKTIARLNLSWLIVHLGVSVMVLPVMEFNPKDVIYFSA
GDFELATATVLAQCLRPVYSLATALFGERNWAFVSVMLPIPTIATNVLHAPGLP
LWPLACALTLGAGGNFAASMTNANAFPHRLKGAAGLGGAGNLGVSVIQVVGML
VIASVGDRLPYIVCGLYVLLII IAGICAMLFMNDIEHHRIGVNTIRPVLVIVSTRDS
WVLALVLAISFGSFGIFGFAFGVLTNFAVGQSTAOAALHAELAFIGPTLAAR
FWGRLADRLGSRVTLVFGAMVFAAGLLGVLTIEGSRVCEIRGVMMASVFAAGFIT
LFTLSLGNIGSVKMLPTTFEACHSHSGINDNCDRMSRVISGVVIGFVAEYALGGV
GIDLALRESYNTGGVTAFWIFMLCYAAAGVLTWKVYVCRPLPGLNLDHEAANAFAS
VGASRTHRG"
/complement(4100..4681)
/genes="ML0845"
/complement(4100..4681)
/genes="ML0845"
/notes="Unknown function. Similar to part of some acyl-CoA
oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2
TR:O65201 (EMBL:AF057043) (692 aa) fasta scores: E():
1.5e-06, 35.2% id in 125 aa. Previously sequenced as

```

```

TR:O32973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,
99.5% id in 193 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC31226.1"
/db_xref="GI:13092925"
/db_xref="SPTREMBL:O32973"
/translation="MPLRPTNAASERVTKETATETILQRMVGARQDSEKKSLFNRCI
QGTFECCECLLSAVARRRQVSKENAFDTEVTQDHLHHAHEHTDRVLETFEL
GIESCNDPAARKLGDIVCDLYALSVKARHGTSSIDTSPSALKAIKAGSRMPMSAAL
CRDVRFRFSQSQSCVTPRCRTGRICPSELQPR"
/complement(4959..5480)
/genes="ML0846"
/pseudo
/complement(4959..5480)
/genes="ML0846"
/notes="Possible pseudogene similar to M. tuberculosis
paralogue Rv1747 (Best blastx score 127)"
/pseudo
/codon_start=1
/transl_table=11
/product="ABC transporter (pseudogene)"
/complement(5522..5848)
/genes="ML0847"
/pseudo
/complement(5522..5848)
/genes="ML0847"
/notes="Possible pseudogene of M. tuberculosis orthologue
Rv2327 (Best blastx score 184)"
/pseudo
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
5906..8080
/genes="ML0848"
5906..8080
/genes="ML0848"
/notes="Similar to Mycobacterium tuberculosis hypothetical
ABC transporter ATP-binding protein Rv2326c or MTCY3G12.08
SW:YN26 MYCTU (P71886) (697 aa) fasta scores: E(): 0,
76.9% id in 697 aa. Shares similar domains with many
ABC-type transporters e.g. Streptomyces roseofulvus ATPase
component of putative ABC transporter frnD TR:068910
(EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%
id in 495 aa and Synecococcus sp. nitrate transport
ATP-binding protein NrtD nrtD SW:NRTD SYN7 (P38046) (274
aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.
Previously sequenced as TR:O32971 (EMBL:Z98741). Contains
hydrophobic, possible membrane-spanning regions. Contains
2 Pfam matches to entry PF00005 ABC_tran, ABC_transporter.
Contains 2 x PS00017 ATP/GTP-binding site motif A
(p-loop). Contains 2 x PS00211 ABC transporters family
signature."
/codon_start=1
/transl_table=11
/product="ABC transporter"
/protein_id="CAC31229.1"
/db_xref="GI:13092926"
/db_xref="GOA:Q9CCF9"
/db_xref="SPTREMBL:Q9CCF9"
/translation="MDLCIHRMALLQCATRCHRWKLTQLCRAAPDLAPGTGTIGKV
TAHPRHLHRSGLSQELAQASVLAALCAVTAIVSVVFPFAAGLALGTVPRVGLLA
YRYRFRULMTANVAGVIAFLITGLGFIATVNSAYIGLTVGVKRGQGLTVIALA
LFAGLGAANVAVVALVLRURLHRTFKAMTANVAGVIAFLITGLGFIATVNSAYIGLTVGVKRGQGLTVIALA
GLOHPWMLLGYFVITILVSLIGMWLVSLRVLRIIPDVHKLDLAFSACNEDAPVGP
VPMWLDKVRFPYFHAGQDALREVSLDRVGEHVAVTGANGSKTTLMLILAGREPTSG
VDRPGAVGLKGLGTAVLQHPESQVLTGVADVVWGLPPGTDVDVNRLLREVGLTD
AFERDTSLSGSGELQRLAALAPSLIADENVSMVDQCRDALLGLVSLGTLKR
HPIALVHTHYNNEADPTNINLSDPDNAGMAETVAPVSTVADVHRHPVLELV
GVGHEYSCTPWAKAALHDISFVRQSDGVLYGNSGSKSTLAWINGLMVPTGAC
LIDGRPTHEHVGAVALSFQAARLQMRSDVLEVASAAGSFREDRVAALGVVGLD

```

Query Match 74.6%; Score 19.4; DB 1; Length 348450;
 Best Local Similarity 95.2%; Pred. No. 4.4e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTA 22
 |||||
 Db 95993 CAACGGCGAGCCCGCAACTA 95973

RESULT 10
 LOCUS AR226916 505 bp DNA linear PAT 20-DEC-2002
 DEFINITION Sequence 376 from patent US 6444799.
 ACCESSION AR226916
 VERSION AR226916.1 GI:27265489
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 505)
 AUTHORS Ross,B.C.
 TITLE P. gingivalis polynucleotides and uses thereof
 JOURNAL Patent: US 6444799-A 376 03-SEP-2002;
 FEATURES Location/Qualifiers
 source 1..505
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 6; Length 505;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 Db 336 CAACGGCATGCGCGCAATCAACGAA 360

RESULT 11
 LOCUS BT010447 1236 bp mRNA linear PLN 26-AUG-2003
 DEFINITION Arabidopsis thaliana At3g02610 mRNA, complete cds.
 ACCESSION BT010447
 VERSION BT010447.1 GI:34222061
 KEYWORDS FLI CDNA
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 1236)
 AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
 Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
 Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
 Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
 Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
 Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
 and Ecker,J.R.

TITLE Arabidopsis ORF clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1236)
 AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
 Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
 Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
 Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
 Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
 Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
 and Ecker,J.R.

TITLE Direct Submission
 JOURNAL Submitted (26-AUG-2003) Salk Institute Genomic Analysis Laboratory
 (SIGAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-Length cDNA'; Seki M., Narusaka M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
 sequenced the PUNI (ORF) clones using the RAFL cDNAs: Shinn, P.,
 Chen, H., Cheuk, R., Kim, C.J., Bowser, L., Chan, M.M., Chang, C.M.,
 Dale, J.M., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J.,
 Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L.,
 Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K.,
 Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.

FEATURES
 source
 Location/Qualifiers
 1..1236
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="U61818"
 /ecotype="Columbia"
 /note="This clone is in pUNI 51"
 1..1236
 /note="acyl-[acyl-carrier-protein]
 desaturase (stearyl-ACP desaturase), putative
 /codon_start=1
 /product="At3g02610"
 /protein_id="AA02867.1"
 /db_xref="GI:34222062"

CDS
 1..1236
 /translation="MALLNSTITVAMKQNPVAVSPRTTCVGSFSPRLRLRVSCV
 ATNPSKTSEETKKKFRPIKEVFNQVTHITTEKLEIFKSMNWAQENLJSLKPV
 SWQPDFELPNTDREDFEQLKELRDPDYPFVVLVDMITTEALPTQYTLNT
 LDGVKDTGGSLTSMVAVRAWTAENRHGDLNKLKLYLSGRVDMRHEKTIQVLI
 GMSKFNPNPYNGPIYTSFOERATFISGNTAKLATYGTTLAKICGTTAADEKRE
 TATYTRIVEKLEIFIDPDGTGVQALASMRKRITMPAHLMDGDDDDLEPHYAAVQRIGV
 TATDYAGILEFLERWEKELGMLSGEGRRAQDYLTLPQIRRLERANDRVKIA
 SKSPSVSFSNVIYGREVEL"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 8; Length 1236;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 Db 339 CAACGACGAGACCGCAATCTACGAA 363

RESULT 12
 LOCUS BT011075 1766 bp mRNA linear INV 19-DEC-2003
 DEFINITION Drosophila melanogaster SD09672 full insert cDNA.
 ACCESSION BT011075
 VERSION BT011075.1 GI:40215457
 KEYWORDS FLI CDNA
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1766)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Achavani,A., Carlson,J.,
 Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
 George,R., Gonzalez,B., Guarin,H., Krommiller,B., Li,P., Liao,G.,
 Miranda,A., Mungall,C.J., Nunco,J., Pacle,J., Paragas,V., Park,S.,
 Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
 and Celniker,S.

TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2003) Berkeley Drosophila Genome Project,

COMMENT
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES
source
1..1766
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1..51766
/genes="SRPK"
/db_xref="FLYBASE:FBgn0026370"
372..>1766
/notes="Longest ORF"
/codon_start=1
/product="SD09672p"
/protein_id="AAR82740.1"
/db_xref="GI:40215458"
/db_xref="FLYBASE:FBgn0026370"
/translation="MNKADVNRVLAIOAKKKRKNRKEKONGTNPQSNSSONON
PRESSNPFEPRNSENLOAPDNANNSHNNAESSAFAFKATATATASRSSGSGAGTPD
QVPPKTKAKPNTQKQHPQPPRSSNESYESSETPSDNEDELMEDYCKGGHPV
NIGDFDRHYHVRKLGWGHFSTWLCWDLQAMGYVAIKIVSAHPFAETARDEKIL
KTREDDPNRRKTVQMLDDPKITGVNGTHICMVFEVLGDNLLKLRKSNYRGIPL
ANVKTITROPLEGLDYLCCKIHTDKPENLVLCVDFPHVSLATATQLYCNWSK
MYPSLVSRAPKEVREBPITGKMSKNRKKLKKKAKRWELFKQORDYLFQADGQGAIN
PREVONGDAGISDAEYFVANSVEDNVHVAQGPSKQREERKAPPEQSEEAQEDPL
TEGTDRAKKKKKKK"

ORIGIN
Query Match 71.5%; Score 18.6; DB 3; Length 1766;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 1561 AGAACGGCGAGGCCGGAATCTCCGA 1585

RESULT 13
AB046101/c
LOCUS
DEFINITION
Macaca fascicularis brain cDNA, clone:Occe-16232.
ACCESSION
AB046101.1 GI:9280278
VERSION
AB046101.1 GI:9280278
KEYWORDS
fis (full insert sequence).
SOURCE
Macaca fascicularis (crab-eating macaque)
ORGANISM
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hitai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA

libraries
Unpublished
2 (bases 1 to 1959)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (14-JUL-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/
tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: macaque brain cDNA library QcCE
Lab host: TOR10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ANGTGGCTTTTTTTTTTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAAGAGCTGCG];
3' end primer [CGACCTGCGCTCGACCA]).
Location/Qualifiers
1..1959
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QcCE-16232"
/clone_lib="macaque brain cDNA library QcCE"
/dev_stage="adult"
1034..1369
/codon_start=1
/product="unnamed protein product"
/protein_id="BAB01683.1"
/db_xref="GI:9280279"
/translation="MLVINTSALNLTFTILPYPENPLFLVLDGLSVPELIGQLL
CSAEVFTSLCNHRSPSPVLSTMTVLGDKTRQLLGNVGVAVGHVCVCTGICRPV
CKLATLLEV"

ORIGIN
Query Match 71.5%; Score 18.6; DB 9; Length 1959;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 1140 ACAACTGAGAGGCCGGAATCTAAGA 1116

RESULT 14
BD180518
LOCUS
DEFINITION
Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION
BD180518.1 GI:30791436
VERSION
BD180518.1 GI:30791436
KEYWORDS
JP 2002325574-A/1009.
Thermus thermophilus
SOURCE
Thermus thermophilus
ORGANISM
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE
1 (bases 1 to 2067)
Kuramitsu, N. and Yokoyama, S.
Highly thermophilic bacterium-derived protein and gene encoding it
Patent: JP 2002325574-A 1009 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
OS
Thermus thermophilus
COMMENT
PN JP 2002325574-A/1009
PD 12-NOV-2002

PF 23-FEB-2001 JP 2001116171
 PI NARUKI KURAMITSU, SHIGEKI YOKOYAMA
 PC C12N15/09, C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, PC
 C12N5/10,
 PC C12N9/88, C12P21/02, (C12N9/88, C12R1:01), (C12N15/09, C12R1:01),
 PC (C12P21/02, C12R1:01), C12N15/00, C12N15/00, C12N5/00, (C12N15/00,
 PC C12R1:01)
 CC Highly thermophilic bacterium-derived protein and gene CC
 encoding it
 FH Key Location/Qualifiers
 FT CDS (1)..(2067)
 FEATURES
 source
 1..2067
 Location/Qualifiers
 /organism="Thermus thermophilus"
 /mol_type="genomic DNA"
 /db_xref="taxon:274"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 6; Length 2067;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 DB 1472 AAAACGGCGAGCGCTGTTCTACGA 1496

QY 1 ACAACGGCGAGCGCGAATCTACGA 25
 |||||
 DB 1472 AAAACGGCGAGCGCTGTTCTACGA 1496

RESULT 15
 AF301149 2368 bp mRNA linear INV 21-FEB-2001
 LOCUS Drosophila melanogaster SR protein kinase 1 mRNA, complete cds.
 DEFINITION
 AF301149
 ACCESSION
 AF301149.1 GI:10242346
 KEYWORDS
 Drosophila melanogaster (fruit fly)
 SOURCE
 Drosophila melanogaster
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 1 (bases 1 to 2368)
 Allemand, E., Gattori, R., Bourbon, H.M., Stevenin, J., Caceres, J.F.,
 Soret, J. and Tazi, J.
 Distinctive features of Drosophila alternative splicing factor RS
 domain: implication for specific phosphorylation, shuttling, and
 splicing activation
 Mol. Cell. Biol. 21 (4), 1345-1359 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 11106380
 REFERENCE
 2 (bases 1 to 2368)
 Allemand, E. and Tazi, J.
 Characterization of the Drosophila ortholog of the human SRPK1
 protein kinase
 Unpublished
 JOURNAL
 REFERENCE
 3 (bases 1 to 2368)
 Allemand, E. and Tazi, J.
 Direct Submission
 TITLE
 Submitted (29-AUG-2000) Centre Nationale de Recherche Scientifique,
 Institut de Genetique Moleculaire de Montpellier UMR-5535, IFR24,
 1919, Route de Mende, Montpellier, Herault 34293, France
 JOURNAL
 FEATURES
 source
 1..2368
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /chromosome="2"
 /map="52A1-52A2"
 1..2295
 /note="SRPK1; similar to Homo sapiens SRPK1"
 /codon_start=1
 /product="SR protein kinase 1"
 /protein_id="AAG15387.1"
 /db_xref="GI:10242347"
 /translation="MNKADVNRVLAIAQKKHKKRKGKNGTNPQGSNQCN"

KPSSNPPEPNSSENLOAPDNANNHLPSSAARSKATATATATASTSGSGSAGTDP
 QIQPKTQAKAKENTQKQQQPPRSSNSESIESIEFSNEDQELMEDICKGGYHPV
 NIGDLFDRHYVIRKLGWGHFSTVWLCWQAMGYVAIKIVKSAPHFAETARDIKIL
 KIVRETDPSNPRRHKTQMLDDFKITGVNGTHICWFEVLGNLLKLRKSNYRGIPL
 ANVTITRQVLEGLDHTCKKIHTDIKPNVLLCVDEPHVRSLEATEATQLYCMNSK
 MVSLSRAPKEYREPTTGMKSNKKKKKKAKRMELFKQORDYLEADGOGAIN
 PNEVQNGAGISDADEVFANDSVNHHVAAQPSRKQORERKAPPEQSEAEQSDPL
 TEGTDKAKKKKKKKKKKSKQSQPPQLENSTSSAESSEALKSQOANGSNSTNNK
 SNTNSGTLKQSGNKKMPLPKPKQKSGHNLNNNSKPNNSDSKISGSEVENTSS
 ATNGPHSNSTLTPPPPPQAKHKKAKKOPALDECNHVKIADLNACWDRHFTDIOT
 RQYRSLEVIIGAGYNTSMTACWVPELATDYLPEPHSGESYTDDEHLAHIIL
 LGPTPEILLNGTYAAKSTFSCELRNLISGLKPMGLMDVLEKYSQKDAASAPSEL
 TPLFEFDNKRATAECLQHPWLR"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 3; Length 2368;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCGCGAATCTACGA 25
 |||||
 DB 1190 AGAACGGCGAGCGCGAATCTCCGA 1214

RESULT 16
 LMFLUNK03/4
 LOCUS LMFLUNK03 40970 bp DNA linear INV 29-OCT-2000
 DEFINITION
 Leishmania major Friedlin cosmid L8530A, unmapped.
 ACCESSION
 AL449123.1 GI:11071675
 VERSION
 AL449123.1
 KEYWORDS
 kinesis-like protein.
 SOURCE
 Leishmania major
 ORGANISM
 Leishmania major
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 REFERENCE
 1 (bases 1 to 40970)
 Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and
 Smith, D.F.
 A physical map of the Leishmania major Friedlin genome
 Genome Res. 8 (2), 135-145 (1998)
 JOURNAL
 MEDLINE
 PUBMED
 98146435
 REFERENCE
 2 (bases 1 to 40970)
 Zimmermann, W., Mambutt, R., Ivens, A.C., Quail, M., Rajandream, M.A.
 and Barrell, B.G.
 Direct Submission
 TITLE
 Submitted (26-OCT-2000) European Leishmania major Friedlin genome
 sequencing project, Sanger Centre, The Wellcome Trust Genome
 Campus, Hinxton, Cambridge CB10 1SA. (E-mail: barrell@sanger.ac.uk)
 and Agowa GmbH, Glienickestr. 185, D-12489, Berlin, Germany
 see <http://www.ebi.ac.uk/parasites/leish.html>
 JOURNAL
 FEATURES
 COMMENT
 Notes:
 Details of leishmania sequencing at the Sanger Centre are available
 on the World Wide Web.
 see <http://www.sanger.ac.uk/Projects/Lmajor/>
 CDS are numbered using the following system eg L8530.01. L8530
 (cosmid name), 01 (first CDS)
 To make the cosmid library Leishmania major Friedlin DNA was
 partially digested with Sau3AI prior to cloning into BamHI site of
 the cosmid shuttle vector cLHYG (Ryan et al. 1993 Gene
 131:145-150). The sequence of the packaged vector was determined by
 Peter Myler and Ken Stuart at Seattle Biomedical Research
 Institute and is available as accession number U59231.
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 Gene prediction is done using:
 (1)
 the Frameplot program of Bibb et al.,
 Gene 30:157-166(1984) as implemented
 at <http://www.nih.gov.jp/~jun/cgi-bin/frameplot.pl>. (2)

codon preference based on the codon usage table for Leishmania at <http://www.kazusa.or.jp/codon/>

(3) the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a kozak sequence. If this cannot be identified we choose the most upstream initiation codon. Transmembrane domains were predicted as implemented at the TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM-1.0/>.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid L8530A was thought to map to chromosome 5, but does not, and is currently not overlapped by any cosmid sequenced to date.

FEATURES

```

source
1..40970
    /organism="Leishmania major"
    /mol_type="genomic DNA"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /clone="cosmid L8530A"
1..174
    /gene="L8530.01"
<1..174
    /gene="L8530.01"
    /note="L8530.01, len > 56 aa, unknown; good similarity to Q9SVU9, hypothetical 22.5 kd protein (206 aa, Arabidopsis thaliana, EMBL: AL161573, CAB81470); Pasta scores: E(1):0.07, 42.0% identity in 50 aa"
    /codon_start=1
    /label="L8530.01"
    /product="conserved hypothetical protein"
    /protein_id="CAC14604.1"
    /db_xref="GI:11071676"
    /db_xref="SPTREMBL:O9GRN5"
    /translation="DRVRSALRSPSLQKLIKIDSRSLDALEAAQYNNADFKGFC DAVRVAEVEGR"
    complement(269..283)
    /note="poly-pyrimidine tract"
789..804
    /note="poly-pyrimidine tract"
792..803
    /note="(cct)4"
908..922
    /note="poly-pyrimidine tract"
1185..1202
    /note="poly-pyrimidine tract"
1273..1294
    /note="poly-pyrimidine tract"
    complement(1490..1506)
    /note="poly-pyrimidine tract"
1507..1522
    /note="(ttg)4"
1727..1814
    /note="region of BLASTN similarity to: 15..102 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 77% identity over 87 bases"
    region of BLASTN similarity to: 18834..18923 AL132764 Leishmania major Friedlin chromosome 4 PAC P1421, 72% identity over 89 bases"
    region of BLASTN similarity to: 29815..29904 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 72% identity over 89 bases"
    region of BLASTN similarity to: 45..112 AL139794
```

```

Leishmania major Friedlin chromosome 4 PAC P1105, 86% identity over 67 bases"
    region of BLASTN similarity to: 27910..27986 AL359781 Leishmania major Friedlin chromosome 21 cosmid L3302, 79% identity over 76 bases"
    region of BLASTN similarity to: 7374..7431 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 84% identity over 57 bases"
    region of BLASTN similarity to: 24835..24899 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 81% identity over 64 bases"
    complement(1733..1801)
    /note="region of BLASTN similarity to: 5998..6066 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 76% identity over 68 bases"
    region of BLASTN similarity to: 2900..2990 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 72% identity over 90 bases"
    region of BLASTN similarity to: 34910..34987 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 86% identity over 57 bases"
    region of BLASTN similarity to: 3605..3700 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 81% identity over 95 bases"
    region of BLASTN similarity to: 284..378 AL161414 Leishmania major Friedlin chromosome 19 cosmid L5515, 80% identity over 94 bases"
    region of BLASTN similarity to: 282..370 AL161414 Leishmania major Friedlin chromosome 19 cosmid L5515, 82% identity over 88 bases"
    region of BLASTN similarity to: 25755..25841 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 82% identity over 86 bases"
    region of BLASTN similarity to: 33949..34050 AL359773 Leishmania major Friedlin chromosome 14 PAC P1295, 73% identity over 101 bases"
    region of BLASTN similarity to: 25749..25840 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 82% identity over 91 bases"
    region of BLASTN similarity to: 3379..3460 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 76% identity over 81 bases"
    region of BLASTN similarity to: 33726..33826 AL359773 Leishmania major Friedlin chromosome 14 PAC P1295, 68% identity over 100 bases"
    region of BLASTN similarity to: 25744..25841 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 80% identity over 97 bases"
    region of BLASTN similarity to: 3591..3701 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 76% identity over 110 bases"
    region of BLASTN similarity to: 24557..24649 AC009605 Leishmania major chromosome 35 clone L8019 strain Friedlin, complete sequence, 73% identity over 92 bases"
    region of BLASTN similarity to: 236017..236106 AF008205 Leishmania major chromosome 1, complete sequence, 75% identity over 89 bases"
    region of BLASTN similarity to: 1548..1649 AL049198 Leishmania major Friedlin chromosome 4 cosmid L5683, 66% identity over 101 bases"
    1860..2342
    /note="region of BLASTN similarity to: 1..483 AQ850336 LMAJFV1 lm35e04.x1 Leishmania major FV1 random genomic library Leishmania major genomic clone LMAJFV1 lm35e04 3, similar to contains element 284-bp.4 Leishmania repetitive element, 99% identity over 482 bases"
    1910..1925
    /note="poly-pyrimidine tract"
    complement(2583..2655)
    /note="region of BLASTN similarity to: 32884..32956 AF239995 Leishmania major strain Friedlin chromosome 5 cosmid L8530, complete sequence, 78% identity over 72 bases"

```

misc_feature

repeat_region

misc_feature

region of BLASTN similarity to: 32885..33003 AF239995
 Leishmania major strain Friedlin chromosome 5 cosmid
 L8530, complete sequence, 68% identity over 118 bases
 region of BLASTN similarity to: 9686..9748 AF239995
 Leishmania major strain Friedlin chromosome 5 cosmid
 L8530, complete sequence, 82% identity over 62 bases
 region of BLASTN similarity to: 22305..22440 AC009605

Query Match 71.5%; Score 18.6; DB 3; Length 40970;

Best Local Similarity 84.0%; Pred. No. 1e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCGAATCTACGAA 26

DB 1349 CACGGCGAGCGCGAATCAACGAA 1325

RESULT 17
 AC017471 AC017471 42950 bp DNA linear HTG 09-DEC-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 ACCESSION AC017471
 VERSION HTG; HTGS PHASE2
 KEYWORDS AC017471.1 GI:6553515
 SOURCE Drosophila melanogaster (fruit fly)

ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 42950)

REFERENCE
 AUTHORS Adams,M. and Venter,J.C.
 TITLE Direct Submission

JOURNAL
 Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10210907 by the submitter.
 For more information on this record e-mail to fly@cslera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..42950
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN

Query Match 71.5%; Score 18.6; DB 2; Length 42950;

Best Local Similarity 84.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCGAATCTACGA 25

DB 10972 AGAACGGCGAGCGCGAATCTCCGA 10996

RESULT 18
 ATAC021640 103904 bp DNA linear PLN 30-OCT-2002
 LOCUS Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence,
 DEFINITION complete sequence.

ACCESSION AC021640

VERSION AC021640.7 GI:12408747

KEYWORDS HTG.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 103904)

REFERENCE
 AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.,
 Wu,D., Rensing,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
 Barnstead,M.B., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence
 Unpublished
 2 (bases 1 to 103904)
 Lin,X. and Kaul,S.

Direct Submission

Submitted (18-JAN-2000) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org

3 (bases 1 to 103904)

Lin,X.

Direct Submission

Submitted (24-JAN-2001) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Jan 24, 2001 this sequence version replaced gi:11280770.

Address all correspondence to:at@tigr.org

BAC clone F16B3 is from Arabidopsis chromosome III and is near the
 molecular marker mi74.

The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov); Genefinder (Phil Green, University of
 Washington); Genscan (Chris Burge, GENSCANW.html), and NetPlantGene
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at/at.html).

Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted as tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Location/Qualifiers
 1..103904

FEATURES

source

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="III"

/map="mi74"

/clone="F16B3"

<129..>979

/gene="F16B3.1"

/note="predicted by genefinder"

join(<129..241,586..>979)

/gene="F16B3.1"

join(129..241,586..979)

/gene="F16B3.1"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAF32445.1"

/db_xref="GI:5957701"

/translation="MEERDWEGLSPSENAGLLCYDDEEYQFGPAPKLPFRYLSEIG
 LQLGDDDDVVISKOLYSGIAEGKCCVENEYVYKLGJYLGRHGVPMWTKY
 AVNTTPSDESLCAAEFFQDRDSVTKLLSDMHICDARPFVDVYLPNSQPKKSGPSP
 SVVTCFSG"

complement(<3079..>4265)

/gene="F16B3.2"

/note="identical to COL2 GB:AB67879 from [Arabidopsis
 thaliana]"

complement(join(<3079..3438,3582..>4265))

/gene="F16B3.2"

complement(join(3079..3438,3582..4265))

/gene="F16B3.2"

/codon_start=1

/product="COL2"

/protein_id="AAF32446.1"

```
/db_xref="GI:6957702"
/translation="MLKEENSGTWARACDTRSAACTVCEADSAYLCTTCDAVH
AANRVAERVRVQCSBPAAFLCADAASLCTCAEIHSAANPLARHORVPII
PLANSACSSAPSETDADDEDDREVASLWLPNGKNTGNQNGFLFGVEYLDLVDYS
SNDNQFENQVTHVORSFGGVVPLQVEESTSLQSQOQFOLGNGPFGSAHYN
NNSLKLNSASVSSMDISVPESATSDITVCHPTTKETIDQSGPQVQOQITPM
REARVLRKRRKRRKDKTIRIARSAKAYAIIRPKRFAIRIETAEAEIEFSTS
LMSGTGIVPSPF"
repeat_region
5228..5255
/db_family="CAA)n"
<5471..>7228
/gene="F16B3.3"
/notes="predicted by genefinder"
<5471..>7228
/gene="F16B3.3"
5471..7228
/gene="F16B3.3"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32447.1"
/db_xref="GI:6957703"
/translation="MVFPFLKLDFTQGRAGDSLGKPGKPGSTIRIGRIVRGNIAIKDA
GTSKLRIVSSENMWIIHDLGSNGTILNSDTIDTPNT-SHGEIKLGYSTILV
NFGSDVQAQPKLRPRRNNKRLAADPDPIESVQEKPKTRGSSKQEEHELPL
KSTRASKNLDDIADKEBELDVEIEKVVKARVGRPKVQAGSAIAKEBEVVEKRVG
RPNKASALTEBEVVVEKGNRARGKNSIYQKSLKLEVEDTPKAVEISEVKS
KVRTRKQENECFGEVDEKRTKSTRKTEIGCESFLEIEMVLNVAARKSRAK
KMDPEKTRNDAGEEVLNKNCHVEEDKNEAQCGSGSDDICQDEKSCDGSK
VEQVEIELRKSTVEGLKTDKEDGETENSQDIERENKEAQCGCSDEKYE
RVGGAKRQVEQIEILRKSTVGEDDLNCTVREDGTENLQIEEEDHEDSDCKVE
AGPATDEBKVGQGNKKVVEVYVLEKMLREMPDAIEVLQPKTIBETEMTEPM
RKSMSVHKHIAEQEKGD"
complement(<7759..>10518)
/gene="F16B3.4"
/notes="predicted by genscan, contains Coesterase domain"
complement(join(<7759..7920,8001..8091,8192..8335,
8796..8861,8964..9101,9182..9283,9433..9507,9638..9718,
10256..>10518))
/gene="F16B3.4"
complement(join(7759..7920,8001..8091,8192..8335,
8796..8861,8964..9101,9182..9283,9433..9507,9638..9718,
10256..>10518))
/gene="F16B3.4"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32448.1"
/db_xref="GI:6957704"
/translation="MQLSPRCRPMSENREANSSEMLLHSGNLSLSPHVRV
SGNSSEGSRIQSQSGFRDIGHAAETILRLSNLLGLDLYIPPTSDGLKP
VYVFTGGAMIGYKAWGLLGLAERDIIVCLDYRNFPQGTISDMVSDAAGLSF
VCNNIAGFGDPNRIYLMGOSAGAHISSCALFEQAIKESRGSISWSVQIKAYFGLS
GGYNLNLVPHFNRLGLYSIFLSIMEGESPKQSPFVRLKDLNVRKAAALPHIL
PHGADYSIPPEASKYFTDALQAEVKAELVYKGTHTDLFLQDPLRGKGLDFDHI
VSMIHADSDALRNDVAPPRKRLVPEFLKLKAGRVSPF"
complement(11800..14456)
/gene="F16B3.5"
complement(join(11800..11972,12355..12394,12527..12640,
12750..12834,12919..13039,13133..13258,13559..13659,
13745..13874,14237..14456))
/gene="F16B3.5"
complement(join(11820..11972,12355..12394,12527..12640,
12750..12834,12919..13039,13133..13258,13559..13659,
13745..13874,14237..14413))
/gene="F16B3.5"
/notes="unknown protein"
/codon_start=1
/protein_id="AAF32449.1"
/db_xref="GI:6957705"
/translation="MAEGEDSQRLLKIAAFAFYENDARWADYWNILIPPHASRP
EVDVHFRKRYQYIDPLVPEMSTSSSSSQARPTATASASASNANEQVRNS
GVSPTSGSATGATSGMRDEQIQFVNAWFIYAVLAVLPIPKNLNRLK
SFWGTACSSLYSLYGRPRANMGLQVYFQSIYAAKDFIYFVCLFVTSHLCLK
PALIPILCRALEQAKELRNFGRSTIYRYLEDPCVWVESNTTILNLSQAEIAG
```

```
FLIISLLSQNRNIIQTFMYWOLLKMLNYQAPVTAGYHQSTWSRIGRTVTPIIQRYAPP
LNPVSAVQRWFR"
<15001..>15660
/gene="F16B3.6"
/notes="predicted by genefinder"
<15001..>15660
/gene="F16B3.6"
15001..15660
/gene="F16B3.6"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32450.1"
/db_xref="GI:6957706"
/translation="MSALRLRNANTPAPELDELSDPTSPSRVLKROMSMSORAMNT
LTSANLSNLLPTGTLIAFOLLTPFTFSGVCHDHTRELTAALLFLAASCQVSSFTD
SVKADDTIYFGVTFKGMWVDVDFSGGLGDLAKIRKRFVNDWIHATLSLVFGAV
ARDKYITDCFPSPSEATKHVLDIVFVGVMCSLLFMVFPARRHGIGILVIGVDVR
R"
complement(<16445..>17870)
/gene="F16B3.7"
/notes="predicted by genscan"
complement(join(<16445..16464,16657..17107,17220..>17870))
/gene="F16B3.7"
complement(join(16445..16464,16657..17107,17220..17870))
/gene="F16B3.7"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32451.1"
/db_xref="GI:6957707"
/translation="MELPATKRLRGLVIFPILITIAPILYFGYLYYSTYKHL
SSSSSAISSPSSYHSSSSSDSYKTEDSPSYDNDYDTHDPKSSLUHNDRLS
ISSNGHQVTPKHEHRRKRRKCDIFSGEWINPKAPYNTTCRAIHQNCIKY
GRPLGFMKWKPKCEDLPDPYELEIVRGTRMAFVGDSVSRNHVQSILCLSRV
EHPGDSQCEENFORWKYKTYETIATFWTHLVRABETGTPGPNKSFYNYLDEPD
PTWASQGEFNYIISGQWFERPLEFDKGRIGCLCYIPGVNVAHFAYRRALR
TTFKILGLENFKEVFLRTFAPSHFEGGEWDKGMQRS"
complement(11673..17778)
/rpt_family="CAT)n"
complement(18048..18074)
Query Match 71.5%; Score 18.6; DB 8; Length 103904;
Best Local Similarity 84.0%; Pred. No. ie+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CAACGGCAGGCGCCGATCTACGAA 26
|||||
Db 72110 CAACGACGACGCGATCTACGAA 72134
|||||
RESULT 19
LMFLCHR31_13
WPCOMMENT
Sequence split into 22 fragments LOCUS LMFLCHR31 Accession AL499621
Fragment Name Begin End
LMFLCHR31_00 1 110000
LMFLCHR31_01 100001 210000
LMFLCHR31_02 200001 310000
LMFLCHR31_03 300001 410000
LMFLCHR31_04 400001 510000
LMFLCHR31_05 500001 610000
LMFLCHR31_06 600001 710000
LMFLCHR31_07 700001 810000
LMFLCHR31_08 800001 910000
LMFLCHR31_09 900001 1010000
LMFLCHR31_10 1000001 1110000
LMFLCHR31_11 1100001 1210000
LMFLCHR31_12 1200001 1310000
LMFLCHR31_13 1300001 1410000
LMFLCHR31_14 1400001 1510000
LMFLCHR31_15 1500001 1610000
LMFLCHR31_16 1600001 1710000
LMFLCHR31_17 1700001 1810000
LMFLCHR31_18 1800001 1910000
```

```

LMFLCHR31_19      1900001    2010000'
LMFLCHR31_20      2000001    2110000
LMFLCHR31_21      2100001    2117963
Continuation (14 of 22) of LMFLCHR31 from base 1300001 (AL499621 Leishmania major chromo-

Query Match          71.5%;   Score 18.6;   DB 2;   Length 110000;
Best Local Similarity 84.0%;   Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CACGGCGGAGGCCGCCTACGAA 26
        ||| ||||| ||||| |||||
Db      24990 CACGGCGGAGGCCGCCTACGAA 25014

RESULT 20      LMFLCHR32_17/c
WPCOMMENT
```

Fragment Name	Begin	End
LMFLCHR32_00	1	110000
LMFLCHR32_01	100001	210000
LMFLCHR32_02	200001	310000
LMFLCHR32_03	300001	410000
LMFLCHR32_04	400001	510000
LMFLCHR32_05	500001	610000
LMFLCHR32_06	600001	710000
LMFLCHR32_07	700001	810000
LMFLCHR32_08	800001	910000
LMFLCHR32_09	900001	1010000
LMFLCHR32_10	1000001	1110000
LMFLCHR32_11	1100001	1210000
LMFLCHR32_12	1200001	1310000
LMFLCHR32_13	1300001	1410000
LMFLCHR32_14	1400001	1510000
LMFLCHR32_15	1500001	1610000
LMFLCHR32_16	1600001	1710000
LMFLCHR32_17	1700001	1810000
LMFLCHR32_18	1800001	1910000
LMFLCHR32_19	1900001	2010000
LMFLCHR32_20	2000001	2110000
LMFLCHR32_21	2100001	2210000
LMFLCHR32_22	2200001	2310000
LMFLCHR32_23	2300001	2410000
LMFLCHR32_24	2400001	2510000
LMFLCHR32_25	2500001	2610000
LMFLCHR32_26	2600001	2710000
LMFLCHR32_27	2700001	2727709

```

Query Match          71.5%; Score 18.6; DB 2; Length 110000;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGGCGAGGCCGGAATCTACGAA 26
        |||||
DDB      80631 CACGGCGAGGGCGCAATCAACGAA 80607
               |||||

RESULT 21
AC008343 LOCUS
DEFINITION Drosophila melanogaster, chromosome 2R, region 51F-52A, BAC clone
AC008343 AC008343.4 GI:13487940
AC008343 HTG.
AC008343 Drosophila melanogaster (fruit fly)
AC008343 Drosophila melanogaster
AC008343 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AC008343 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
AC008343 Ephydroidea; Drosophilidae; Drosophila.
AC008343 1 (bases 1 to 158402)
REFERENCE

```

AUTHORS	Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferrier, S., Frise, E., Galle, R.F., Gar, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattai, B., Moshrefi, A., McKittrick, T.C., Moy, M., Murphy, B., Nelson, C., Neilson, K.A., Nuncio, J., Paclet, J., Paragov, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE	Sequencing of Drosophila chromosome 2R, region 51F-52A
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 158402) Celniker, S.E., Agbavanyi, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT	On Mar 29, 2001 this sequence version replaced gi:6563415. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (http://www.fruitfly.org/sequence/) or send email to bdgpf@fruitfly.berkeley.edu .

```

Location/Qualifiers
1. 158402
  /organism="Drosophila melanogaster"
  /mol_type="genomic DNA"
  /strain="y; cn bw sp"
  /db_xref="taxon:7227"
  /chromosome="2R"
  /map="51F-52A"
  /clones="BACR03D24 (D931)"
  /clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBace3.6)"

Query Match 71.5%; Score 18.6; DB 3; Length 158402;
Best Local Similarity 84.0%; Pred. No.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

y 1 ACAACGGCGAGCCCGAATCTACGA 25
  ||||| ||||| ||||| |||||
b 34691 AGAACGGCGAGCCCGAATCTCCGA 34715

RESULT 22
C011127/c
OCUS AC011127 164840 bp DNA linear HTG 13-MAY-2001
Homo sapiens chromosome 18 clone RP11-110J22 map 18, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
CESSION AC011127.6 GI:11136813
VERSION AC011127.6
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 164840)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone RP11-110J22
Unpublished
2 (bases 1 to 164840)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, K., Kann, L., Karatas, A., Klein, J.,
Lehocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, I., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Seaver, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 10, 2000 this sequence version replaced gi:8072498.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research.
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2301
Center clone name: 110 J_22
----- Summary Statistics
Sequencing vector: M13; M77815; 43% of reads
Sequencing vector: Plasmid; n/a; 57% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159887 bases at least Q40
Consensus quality: 182253 bases at least Q30
Consensus quality: 163278 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 163840; sum-of-contigs
Quality coverage: 7.2 in Q20 bases; agarose-fp
Quality coverage: 7.2 in Q20 b.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2679: contig of 2679 bp in length
* 2680 2779: gap of 100 bp
* 2780 4151: contig of 1372 bp in length
* 4152 4251: gap of 100 bp
* 4252 5588: contig of 1337 bp in length
* 5589 5688: gap of 100 bp
* 5689 7014: contig of 1326 bp in length
* 7015 7114: gap of 100 bp
* 7115 15534: contig of 8420 bp in length
* 15535 15634: gap of 100 bp
* 15635 21293: contig of 5659 bp in length
* 21294 21393: gap of 100 bp
* 21394 32481: contig of 11088 bp in length
* 32482 32581: gap of 100 bp
* 32582 45755: contig of 13174 bp in length
* 45756 45855: gap of 100 bp
* 45856 64019: contig of 18164 bp in length
* 64020 64119: gap of 100 bp

Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McNeethers, R., McNeill, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivaz, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
Tirell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission
Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 169553)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, M., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicoll, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teefaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 169553)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicoll, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 17, 2002 this sequence version replaced gi:21307364.

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L3356

Center clone name: 13_L_22

FEATURES

Source	Location/Qualifiers
1..169553	/organism="Homo sapiens"
	/mol_type="Genomic DNA"
	/db_xref="taxon:9606"
	/clone="RP11-13L22"
	/clone_lib="RPC1-11 Human Male BAC"
repeat_region	complement(1..119)
	/rpt_family="L1"
repeat_region	1105..1204
	/rpt_family="MERSA"
repeat_region	1263..1554
	/rpt_family="AluX"
repeat_region	1699..2743
	/rpt_family="L1ME1"
repeat_region	2744..2765
	/rpt_family="(TA)n"
repeat_region	complement(2784..3064)
	/rpt_family="AluX"
repeat_region	complement(3067..3364)
	/rpt_family="AluJ"
repeat_region	complement(3372..3524)
	/rpt_family="MER3"
repeat_region	complement(3834..4135)
	/rpt_family="AluX"
repeat_region	complement(4141..4238)
	/rpt_family="FRAM"
repeat_region	complement(4266..4569)
	/rpt_family="AluY"
repeat_region	4570..4594
	/rpt_family="(T)n"
repeat_region	complement(4632..4933)
	/rpt_family="AluJb"
repeat_region	complement(4940..5051)
	/rpt_family="Charlie"
repeat_region	524..6573
	/rpt_family="L2"
repeat_region	6829..7132
	/rpt_family="AluX"
repeat_region	complement(7400..7478)
	/rpt_family="L1ME3"
repeat_region	complement(7479..7777)
	/rpt_family="AluX"
repeat_region	complement(7778..7906)
	/rpt_family="L1ME3"
repeat_region	8037..8217
	/rpt_family="L2"
repeat_region	complement(8556..8624)
	/rpt_family="AluJ/FRAM"
repeat_region	complement(8689..8732)
	/rpt_family="L1ME3"
repeat_region	complement(8733..9140)
	/rpt_family="L1ME2"
repeat_region	complement(9141..9515)
	/rpt_family="L1ME3"
repeat_region	complement(9646..9713)
	/rpt_family="MLT1J2"
repeat_region	complement(9809..9878)
	/rpt_family="MLT1J"
repeat_region	complement(10743..11631)
	/rpt_family="L2"
repeat_region	11632..11937
	/rpt_family="AluX"
repeat_region	complement(11938..12833)
	/rpt_family="L2"
repeat_region	12834..13128
	/rpt_family="AluJ"
repeat_region	complement(13129..13500)
	/rpt_family="L2"
repeat_region	complement(13555..13705)
	/rpt_family="L2"
repeat_region	13706..14020
	/rpt_family="AluX"

```

repeat_region complement(14021..14065)
/rpt_family="L2"
repeat_region 14070..14121
/rpt_family="(CGAGAA)n"
repeat_region complement(14204..14259)
/rpt_family="MIR"
repeat_region complement(14599..14911)
/rpt_family="AluSq"
repeat_region 16196..16231
/rpt_family="(GGA)n"
repeat_region 16243..16315
/rpt_family="Alu/FLAM"
repeat_region complement(16561..17000)
/rpt_family="MT1C"
repeat_region complement(17059..17174)
/rpt_family="MT1C"
repeat_region complement(17180..17236)
/rpt_family="MT1C"
repeat_region 18051..18366
/rpt_family="AluSp"
repeat_region complement(18447..18749)
/rpt_family="AluSq"
repeat_region complement(19254..19367)
/rpt_family="MIR"
repeat_region 22392..22413
/rpt_family="AT-rich"
repeat_region complement(22681..22989)
/rpt_family="LIME4A"
repeat_region 22990..23295
/rpt_family="AluSx"
repeat_region complement(23296..23329)
/rpt_family="LIME4A"

```

Query Match 71.5%; Score 18.6; DB 9; Length 169553;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 ACAACGCGAGCGCCGAATCTACA 25
|||||
Db 163397 ACAACTGAGAGCGCCGAATCTAAGA 163373

```

```

RESULT 24
AC023532/c
LOCUS 189412 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-150F7 map 11, WORKING DRAFT
SEQUENCE 39 unordered pieces.
ACCESSION AC023532
VERSION AC023532.5 GI:10047802
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 189412)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bedalov, F., Boguslavskiy, L.,
Boukhalil, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenstermaker, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Laroque, K., Lechoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
McEwan, P., McCurk, A., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, F.M.,

```

Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE JOURNAL

COMMENT

Direct Submission
 Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 10, 2000 this sequence version replaced gi:7657721.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5060

Center clone name: 150 F 7

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 15822 bases at least Q40

Consensus quality: 16815 bases at least Q30

Consensus quality: 17267 bases at least Q20

Insert size: 178000; agarose-fp

Insert size: 185612; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 4164: contig of 4164 bp in length
* 4165: gap of 100 bp
* 4265: contig of 1071 bp in length
* 5336: gap of 100 bp
* 5435: gap of 100 bp
* 5436: contig of 1130 bp in length
* 6565: gap of 100 bp
* 6666: gap of 100 bp
* 7843: contig of 1178 bp in length
* 7943: gap of 100 bp
* 7944: gap of 100 bp
* 9118: contig of 1175 bp in length
* 9218: gap of 100 bp
* 9219: gap of 100 bp
* 10375: contig of 1157 bp in length
* 10376: gap of 100 bp
* 10475: gap of 100 bp
* 11651: contig of 1176 bp in length
* 11652: gap of 100 bp
* 11752: contig of 1305 bp in length
* 13057: gap of 100 bp
* 13157: contig of 1183 bp in length
* 13157: gap of 100 bp
* 13430: gap of 100 bp
* 14440: contig of 1294 bp in length
* 15734: gap of 100 bp
* 15834: contig of 2056 bp in length
* 17889: gap of 100 bp
* 17890: contig of 1056 bp in length
* 17990: gap of 100 bp
* 19046: contig of 1435 bp in length
* 19145: gap of 100 bp
* 20580: contig of 1435 bp in length
* 20581: gap of 100 bp
* 20680: gap of 100 bp
* 22492: contig of 1812 bp in length
* 22592: gap of 100 bp
* 24013: contig of 1421 bp in length
* 24113: gap of 100 bp
* 25801: contig of 1688 bp in length
* 24114

```



```

* 25802 25901: gap of 100 bp
* 25902 27707: contig of 1806 bp in length
* 27708 27807: gap of 100 bp
* 27808 30368: contig of 2561 bp in length
* 30369 30468: gap of 100 bp
* 30469 32511: contig of 2043 bp in length
* 32512 32612: gap of 100 bp
* 32612 34829: contig of 2218 bp in length
* 34830 34929: gap of 100 bp
* 34930 37058: contig of 2129 bp in length
* 37059 37159: gap of 100 bp
* 37159 39857: contig of 2699 bp in length
* 39858 39957: gap of 100 bp
* 39958 42581: contig of 2623 bp in length
* 42581 42680: gap of 100 bp
* 42680 46613: contig of 3933 bp in length
* 46614 46713: gap of 100 bp
* 46714 49816: contig of 3103 bp in length
* 49817 49916: gap of 100 bp
* 49917 52697: contig of 2781 bp in length
* 52698 52797: gap of 100 bp
* 52798 55438: contig of 2841 bp in length
* 55439 55539: gap of 100 bp
* 55539 96844: contig of 41306 bp in length
* 96845 96944: gap of 100 bp
* 96945 101635: contig of 4691 bp in length
* 101636 101735: gap of 100 bp
* 101736 106505: contig of 4770 bp in length
* 106506 108605: gap of 100 bp
* 108606 112283: contig of 5678 bp in length
* 112284 112383: gap of 100 bp
* 112384 119345: contig of 6962 bp in length
* 119346 119445: gap of 100 bp
* 119446 127125: contig of 7680 bp in length
* 127126 127225: gap of 100 bp
* 127226 136592: contig of 9367 bp in length
* 136593 136692: gap of 100 bp
* 136693 147694: contig of 11002 bp in length
* 147695 147794: gap of 100 bp
* 147795 159758: contig of 11964 bp in length
* 159759 159858: gap of 100 bp
* 159859 173806: contig of 13948 bp in length
* 173807 173906: gap of 100 bp
* 173907 185740: contig of 11834 bp in length
* 185741 185841: gap of 100 bp
* 185841 189412: contig of 3572 bp in length.
FEATURES
    Source
    1. .189412
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="11"
        /map="11"
        /clones="RP11-150F7"
        /clone.lib="RPC1-11 Human Male BAC"
    1. .4164
        /note="assembly_fragment"
        /clone_end:SP6
        vector_side:left
        4265. .5335
        /note="assembly_fragment"
        5436. .6565
        /note="assembly_fragment"
        6666. .7843
        /note="assembly_fragment"
        7944. .9118
        /note="assembly_fragment"
        9219. .10375
        /note="assembly_fragment"
        10476. .11651
        /note="assembly_fragment"
        11752. .13056
        /note="assembly_fragment"

```

```

misc_feature 13157. .14339
    /note="assembly_fragment"
misc_feature 14440. .15733
    /note="assembly_fragment"
misc_feature 15834. .17889
    /note="assembly_fragment"
misc_feature 17990. .19045
    /note="assembly_fragment"
misc_feature 19146. .20580
    /note="assembly_fragment"
misc_feature 20681. .22492
    /note="assembly_fragment"
misc_feature 22593. .24013
    /note="assembly_fragment"
misc_feature 24114. .25801
    /note="assembly_fragment"
misc_feature 25902. .27707
    /note="assembly_fragment"
misc_feature 27808. .30368
    /note="assembly_fragment"
misc_feature 30469. .32511
    /note="assembly_fragment"
misc_feature 32612. .34829
    /note="assembly_fragment"
misc_feature 34930. .37058
    /note="assembly_fragment"

Query Match 71.5%; Score 18.6; DB 2; Length 189412;
Best Local Similarity 84.0%; Pred. No. le+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGATCTACGAA 26
    |||||
Db 181908 CAACAGCGAGCGCCGCTCTACAA 181884

RESULT 25
AP001397
LOCUS 197837 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-825H23 map 18q22, WORKING
DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION AP001397
VERSION AP001397.3 GI:9188493
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 197837)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 197,837 genomic DNA of 18q22
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 197837)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8117308.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-825H23
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads

```


Assembly program: Phrap; version 0.990329
Consensus quality: 190215 bases at least Q40
Consensus quality: 193565 bases at least Q30
Consensus quality: 194985 bases at least Q20
Insert size: 195937; sum-of-contigs
Quality coverage: 9.11x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 22006 contig of 22006 bp in length
22107 42860 contig of 20754 bp in length
42961 63463 contig of 20503 bp in length
63564 81431 contig of 17868 bp in length
81532 99656 contig of 18125 bp in length
99757 116826 contig of 17070 bp in length
116927 132183 contig of 15257 bp in length
132284 142638 contig of 10355 bp in length
142739 152508 contig of 9770 bp in length
152609 161840 contig of 9232 bp in length
161941 169129 contig of 7189 bp in length
169230 175273 contig of 6044 bp in length
175374 180468 contig of 5095 bp in length
180569 183944 contig of 3376 bp in length
184045 187766 contig of 3722 bp in length
187867 190600 contig of 2734 bp in length
190701 193108 contig of 2307 bp in length
193108 194983 contig of 1876 bp in length
195084 196452 contig of 1369 bp in length
196553 197837 contig of 1285 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
1 22006: contig of 22006 bp in length
22007 22106: gap of 100 bp
42860: contig of 20754 bp in length
42961 42960: gap of 100 bp
63463: contig of 20503 bp in length
63564 63563: gap of 100 bp
81431: contig of 17868 bp in length
81532 81531: gap of 100 bp
99656: contig of 18125 bp in length
99757 99756: gap of 100 bp
116826: contig of 17070 bp in length
116927 116926: gap of 100 bp
132183 132182: contig of 15257 bp in length
132284 132283: gap of 100 bp
142638: contig of 10355 bp in length
142739 142738: gap of 100 bp
152508: contig of 9770 bp in length
152609 152608: gap of 100 bp
161840 161839: contig of 9232 bp in length
161941 161940: gap of 100 bp
169129 169128: contig of 7189 bp in length
169230 169229: gap of 100 bp
175273 175272: contig of 6044 bp in length
175374 175373: gap of 100 bp
180468 180467: contig of 5095 bp in length
180569 180568: gap of 100 bp
183944 183943: contig of 3376 bp in length
184045 184044: gap of 100 bp
187766 187765: contig of 3722 bp in length
187867 187866: gap of 100 bp
190600: contig of 2734 bp in length
```

```
* 190601 190700: gap of 100 bp
* 190701 193007: contig of 2307 bp in length
* 193008 193107: gap of 100 bp
* 193108 194983: contig of 1876 bp in length
* 194984 195083: gap of 100 bp
* 195084 196452: contig of 1369 bp in length
* 196453 196552: gap of 100 bp
* 196553 197837: contig of 1285 bp in length.

FEATURES
    source
        1. .197837
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="18"
            /map="18q22"
            /clone="RP11-825H23"
            /note="assembly_fragment clone_end:SP6 vector_side:left"
        misc_feature
            1. .22006
                /note="assembly_fragment"
            misc_feature
                22107. 42860
                    /note="assembly_fragment"
            misc_feature
                42961. 63463
                    /note="assembly_fragment"
            misc_feature
                63564. 81431
                    /note="assembly_fragment"
            misc_feature
                81532. 99656
                    /note="assembly_fragment"
            misc_feature
                99757. 116826
                    /note="assembly_fragment"
            misc_feature
                116927. 132183
                    /note="assembly_fragment"
            misc_feature
                132284. 142638
                    /note="assembly_fragment"
            misc_feature
                142739. 152508
                    /note="assembly_fragment"
            misc_feature
                152609. 161840
                    /note="assembly_fragment"
            misc_feature
                161941. 169129
                    /note="assembly_fragment"
            misc_feature
                169230. 175273
                    /note="assembly_fragment"
            misc_feature
                175374. 180468
                    /note="assembly_fragment"
            misc_feature
                180569. 183944
                    /note="assembly_fragment clone_end:T7 vector_side:left"
            misc_feature
                184045. 187766
                    /note="assembly_fragment"
            misc_feature
                187867. 190600
                    /note="assembly_fragment"
            misc_feature
                190701. 193007
                    /note="assembly_fragment"
            misc_feature
                193108. 194983
                    /note="assembly_fragment"
            misc_feature
                195084. 196452
                    /note="assembly_fragment"
            misc_feature
                196553. 197837
                    /note="assembly_fragment"

ORIGIN
    Query Match          71.5%; Score 18.6; DB 2; Length 197837;
    Best Local Similarity 84.0%; Pred No. 1e+03;
    Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

    QY 1 ACAACGGGAGGCCCGAATCTACGA 25
        |||||
    Db 125189 ACAACTGAGAGCCGAGATCTAAGA 125213

RESULT 26
AP002985/c 205152 bp DNA linear PRI 15-MAR-2003
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:CTD-2542M24,
DEFINITION complete sequence.
ACCESSION AP002985
```

```

VERSION      AP002985.2  GI:16751492
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens genomic DNA
JOURNAL      Published Only in Database (2000)
REFERENCE    2 (bases 1 to 205152)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Sushiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      On Nov 5, 2001 this sequence version replaced gi:11559300.
FEATURES     source
             1..205152
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="11"
                /map="11q"
                /clone="CTD-2542M24"
ORIGIN
Query Match      71.5%; Score 18.6; DB 9; Length 205152;
Best Local Similarity 84.0%; Pred.NO.le+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGGCAGGCGCCGAACTACGAA 26
      ||||| ||||| ||||| ||||| |||||
Db      50297 CAACAGCGAGGCGCCGCTCTACAA 50273

RESULT 27
AP001447
LOCUS      Homo sapiens chromosome 11 clone RP11-705P3 map 11q14, WORKING
DEFINITION DRAFT SEQUENCE, 46 unordered pieces.
ACCESSION  AP001447.2  GI:8117321
VERSION     HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 214701)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 214,701 genomic DNA of 11q14
JOURNAL     Published Only in Database (2000)
REFERENCE   2 (bases 1 to 214701)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (15-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@sc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7262580.
            ----- Genome Center
            Center: RIKEN Genomic Sciences Center (GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@sc.riken.go.jp

```

```

----- Project Information
Center project name: HumDraft11
Center clone name: RP11-705P3
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 182011 bases at least Q40
Consensus quality: 197127 bases at least Q30
Consensus quality: 205118 bases at least Q20
Insert size: 210201; sum-of-contigs
Quality coverage: 4.34x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 14737 contig of 14737 bp in length
14838 28083 contig of 13246 bp in length
28184 40511 contig of 12328 bp in length
40612 51898 contig of 11287 bp in length
51999 64081 contig of 12083 bp in length
64182 74827 contig of 10646 bp in length
74928 84831 contig of 9904 bp in length
84932 91901 contig of 6969 bp in length
92001 99014 contig of 7014 bp in length
99115 105728 contig of 6614 bp in length
105829 110842 contig of 5014 bp in length
110943 116467 contig of 5525 bp in length
116568 121590 contig of 5023 bp in length
121691 127550 contig of 5860 bp in length
127651 132127 contig of 4477 bp in length
132228 136727 contig of 4500 bp in length
136828 141378 contig of 4551 bp in length
141479 145693 contig of 4215 bp in length
145794 149654 contig of 3861 bp in length
149755 154794 contig of 5040 bp in length
154895 158481 contig of 3587 bp in length
158582 164776 contig of 6195 bp in length
164877 171761 contig of 2885 bp in length
167862 171113 contig of 3252 bp in length
171214 174080 contig of 2867 bp in length
174181 177056 contig of 2876 bp in length
177157 179451 contig of 2295 bp in length
179552 182040 contig of 2489 bp in length
182141 184418 contig of 2278 bp in length
184519 185523 contig of 2005 bp in length
185624 188462 contig of 1839 bp in length
188563 190196 contig of 1634 bp in length
190297 192401 contig of 2105 bp in length
192502 194597 contig of 2096 bp in length
194698 198254 contig of 1557 bp in length
198355 198593 contig of 2239 bp in length
198694 200646 contig of 1953 bp in length
200747 202409 contig of 1663 bp in length
202510 204484 contig of 1975 bp in length
204585 206037 contig of 1453 bp in length
206138 207948 contig of 1811 bp in length
208049 209677 contig of 1629 bp in length
209778 210908 contig of 1131 bp in length
211009 212394 contig of 1386 bp in length
212495 213510 contig of 1016 bp in length
213611 214701 contig of 1091 bp in length

```

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.
1
14737: contig of 14737 bp in length
14738: gap of 100 bp
14838: contig of 13246 bp in length
28084: contig of 100 bp
28184: contig of 12328 bp in length
40511: gap of 100 bp
40611: gap of 100 bp
51888: contig of 11287 bp in length
51899: gap of 100 bp
64081: contig of 12083 bp in length
64082: gap of 100 bp
64182: contig of 10646 bp in length
74828: gap of 100 bp
74927: gap of 100 bp
84831: contig of 9904 bp in length
84931: gap of 100 bp
91900: contig of 6969 bp in length
92000: gap of 100 bp
92001: contig of 7014 bp in length
99014: gap of 100 bp
99114: gap of 100 bp
105728: contig of 6614 bp in length
105729: gap of 100 bp
110842: contig of 5014 bp in length
110843: gap of 100 bp
110943: contig of 5525 bp in length
116567: gap of 100 bp
116568: contig of 5023 bp in length
121590: gap of 100 bp
121591: gap of 100 bp
127550: contig of 5860 bp in length
127551: gap of 100 bp
127552: gap of 100 bp
132127: contig of 4477 bp in length
132128: gap of 100 bp
132228: contig of 4500 bp in length
136727: gap of 100 bp
136728: gap of 100 bp
141378: contig of 4551 bp in length
141379: gap of 100 bp
145693: contig of 4215 bp in length
145694: gap of 100 bp
145794: contig of 3861 bp in length
149754: gap of 100 bp
149755: contig of 5040 bp in length
154795: gap of 100 bp
154895: contig of 3587 bp in length
158481: gap of 100 bp
158482: contig of 6195 bp in length
164776: gap of 100 bp
164777: contig of 2885 bp in length
167761: gap of 100 bp
171113: contig of 3252 bp in length
171114: gap of 100 bp
171214: contig of 2867 bp in length
174080: gap of 100 bp
174081: contig of 2876 bp in length
177056: gap of 100 bp
177157: contig of 2295 bp in length
179451: gap of 100 bp
179452: contig of 2489 bp in length
182040: gap of 100 bp
182041: contig of 2278 bp in length
184148: gap of 100 bp
184518: contig of 2005 bp in length
184519: gap of 100 bp
186523: contig of 1839 bp in length
186524: gap of 100 bp
188463: contig of 1634 bp in length
190196: gap of 100 bp
190197: contig of 2105 bp in length
192401: gap of 100 bp
192402: contig of 2096 bp in length
194597: gap of 100 bp
194598: contig of 1557 bp in length
196254: gap of 100 bp
196255: contig of 2239 bp in length
196355: gap of 100 bp
198593: contig of 2239 bp in length

* 198594 198693: gap of 100 bp
* 198694 200646: contig of 1353 bp in length
* 200647 200746: gap of 100 bp
* 200747 202409: contig of 1663 bp in length
* 202410 202509: gap of 100 bp
* 202510 204484: contig of 1375 bp in length
* 204485 204584: gap of 100 bp
* 204585 206037: contig of 1453 bp in length
* 206038 206137: gap of 100 bp
* 206138 207948: contig of 1811 bp in length
* 207949 208048: gap of 100 bp
* 208049 209677: contig of 1629 bp in length
* 209678 209777: gap of 100 bp
* 209778 210908: contig of 1131 bp in length
* 210909 211008: gap of 100 bp
* 211009 212394: contig of 1386 bp in length
* 212395 212494: gap of 100 bp
* 212495 213510: contig of 1016 bp in length
* 213511 213610: gap of 100 bp
* 213611 214701: contig of 1091 bp in length.
FEATURES
Location/Qualifiers
source
1..214701
/organism="Homo sapiens"
Query Match 71.5%; Score 18.6; DB 2; Length 214701;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CAACGCGAGCGCCGATCTACGAA 26
DB 137150 CAACGCGAGCGCCGCTCTACAAA 137174
RESULT 28
AE003811
LOCUS
DEFINITION
Drosophila melanogaster chromosome 2R, section 39 of 74 of the complete sequence.
ACCESSION
AE003811.3 GI:21627138
VERSION
AE003811.3
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1. (bases 1 to 255652)
AUTHORS
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazer V.J., Blazer M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Cabor G.L., Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,

Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,P., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodgett,
Worley,K.C., Wu,Q., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,B.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
10731132
2 (bases 1 to 255652)
Celniker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A.,
Evans,C.A., Cochrane,J.D., Amanatides,P.G., Brandon,R.C., Rogers,V.,
Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegaw,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D.,
Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
Sequencing of *Drosophila melanogaster* genome
Unpublished
3 (bases 1 to 255652)
Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
Campbell,K., Hradscky,P., Huang,Y., Kaninker,J.S., Prochnik,S.E.,
Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W.,
Celniker,S.E., Clump,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B.,
Milburn,G.H., Richter,J., Russo,S., Seale,S.M.J., Smith,E.,
Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,
Rubin,G.M., Mungall,C.J. and Lewis,S.E.
Annotation of *Drosophila melanogaster* genome
Unpublished
4 (bases 1 to 255652)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
5 (bases 1 to 255652)
FlyBase
Direct Submission
Submitted (06-SEP-2002) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
6 (bases 1 to 255652)
FlyBase
Direct Submission
Submitted (23-JAN-2003) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
On Jun 28, 2002 this sequence version replaced gi:10727530.
Location/Qualifiers
1. 255652
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="2R"
/note="genotype: y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
1884. .36739
/gene="igl"
/locus_tag="CG18285"
/note="last curated on Tue Feb 26 11:36:14 PST 2002"
/map="51E3-51E5"
/_db_xref="FLYBASE:FBgn0013467"

Mrna

Join(1884. .2660,26958. .27155,35201. .36739)
/gene="igl"
/locus_tag="CG18285"
/product="CG18285-PA"
/db_xref="FLYBASE:FBgn0013467"
Join(2363. .2860,26958. .27155,35201. .35427)
/gene="igl"
/locus_tag="CG18285"
/note="igl gene product from transcript CG18285-RA"
/codon_start=1
/product="CG18285-PA"
/protein_id="AAF58168.2"
/db_xref="GI:21627139"
/translation="MGCNTSQELTKDGAAMDVANGPEPSPAPPLLEGESKSSASNH
TNHAKSSIIISNGEAKANGGAVGGSGKSEATNGIDRCDKAAITFEFDDDEAKA
ATKIQAVFRGHKVTWKTKSETKTATNGSAGAAAPSAASAAASAEPTKALEAEF
DPNDKDLCHALKIQSTFRGLARKLVNKDAPEDEIQITKKVAELDIDLDTPELN
KAAKIQASFRGHKTRKDAPE"
Join(11362. .11545,26958. .27155,35201. .36739)
/gene="igl"
/locus_tag="CG18285"
/product="CG18285-RB"
/db_xref="FLYBASE:FBgn0013467"
Join(27014. .27155,35201. .35427)
/gene="igl"
/locus_tag="CG18285"
/note="igl gene product from transcript CG18285-RB"
/codon_start=1
/product="CG18285-PB"
/protein_id="AAM68523.1"
/db_xref="GI:21627140"
/translation="MKKSETKTANNSSAGAAPSAASAAASAEPTKALEAEFP
NDKDLCHALKIQSTFRGLARKLVNKDAPEDEIQITKKVAELDIDLDTPELNKA
ATKIQASFRGHKTRKDAPE"
complement(45752. .47622)
/locus_tag="CG8090"
/note="last curated on Wed Feb 27 13:47:00 PST 2002"
/map="51E5-51E5"
/_db_xref="FLYBASE:FBgn0033995"
complement(Join(45752. .46027,46080. .46265,46323. .46504,
46564. .46748,46816. .47041,47102. .47372,47430. .47622))
/locus_tag="CG8090"
/_db_xref="FLYBASE:FBgn0033995"
complement(Join(45752. .46027,46080. .46265,46323. .46504,
46564. .46748,46816. .47041,47102. .47372))
/locus_tag="CG8090"
/note="CG8090 gene product"
/codon_start=1
/product="CG8090-PA"
/protein_id="AAF58167.1"
/_db_xref="GI:7303100"
/_db_xref="FLYBASE:FBgn0033995"
/translation="MIFAGWLFNKLKFKHYEIRHISVOLIFPSFTFALSITWFELI
IFEIDLVLESSRYFHWRLGLTLLFTMTAVIPIYICSVIHSIFSFSDKWRILTF
MSYFIKPVSRNDIICFERRALITVEMLSAKKRIAMAIYNNKLNPSKPIWMDLASA
VORNTSGEDINOLKQVYGCIFELLRSVLELSLKNMERQWSOTLKPKGFNVLGH
FVSVCYKIFMCCINIIFRVGRKDPVTRGLEIAIHWCGNIDFAFNQHSIFLVGH
CIVTSIRGLLLTLTKFIRISSKSSNIIVILGQINGMVFYFSSVLLMRNMPAEYR
VIITVLGNLHNFYHFWDFVFLVSALTITIIIVLSRKPKVRVDDSDLN"
47969. .50918
/locus_tag="CG11807"
/note="last curated on Tue Feb 26 11:38:34 PST 2002"
/map="51E5-51E5"
/_db_xref="FLYBASE:FBgn0033996"
Join(47969. .48276,48339. .48572,49467. .50346,50406. .50918)
/locus_tag="CG11807"
/product="CG11807-RA"
/_db_xref="FLYBASE:FBgn0033996"
Join(48061. .48276,48339. .48572,49467. .50346,50406. .50551)

gene

Source

/transl_table=11
/product="DNA-binding response regulator, LuxR family"
/protein_id="AA066484.1"
/db_xref="GI:34397422"
/translation="MISIVLVDDHLYRVGLRAAISOMQGLAIVGECSCSELETFLL
KEQEPGLIILDPGNGVNIARKLKLPAVKIILISSEVSEKTVNELLAIDVGG
YICKAQINDIGNAIRVTVTGSHFYGKSVKMLDLCVRSADQSHKKKQQQQND
SSBLTOREKVOYVCLDGYSAKETAENKMLSYRVTETHRNILHLKLGFTNAABELIRYA
VKSGLVDQWQ"
complement(5785..7650)
/locus_tag="PGI432"
complement(5785..7650)
/locus_tag="PGI432"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AA066485.1"
/db_xref="GI:34397423"
/translation="MFLTKNYYCSROIQDTIMINIHTRLTILTLFLGCVFMLSKEIN
VDSLFQIADRTLSATKKRYEEMKQELHQOVSADAYSIAIKFCRDGGSTEEEL
PFLIKYATMATHAGRYSAIMALEEIMRLIKNDNIIILARTYMQIGVFFFGSTEE
DALVYERALOIAOKKEERGISTAYNNMANIYOKNGNTQOEAISYNNKALDIQRTGD
SVSCNSIMGTMTINEKNPDSYKPLKEALTASKINDIEIQALSAYHALYVYAWR
SEYKANEVLTIENACHAGYNQTRLAIDIASIYAEVGLYSLAVECOKRKLAD
SIAKQOEKITEFDIRPKSKBAEIALKNQELHARKLHIALIILSTLLVATITL
IIHAARYQRKLRWMDTNRLLSIIISHDIPGPAQKALQWALDITLSSSDCGSMTR
NMLSIRNSVATELTLLQNLDDWSNIQIGNLAPFVFNIVENIRVINEIYSARNK
DISLHLEAPQCLVHADQOMINTVIRNLLSNAIKFSKAGSVIHRVVEEKAGVIGKVS
VSDGIGNSKKQITDILESNDKKNKATAGTDEIGSLGILICKLLKKQKQVMH"
7932..8147
/locus_tag="PGI433"
7932..8147
/locus_tag="PGI433"
/note="Identified by match to TIGR protein family HMM
TIGR01280"
/codon_start=1
/transl_table=11
/product="hydrolase"
/protein_id="AA066486.1"
/db_xref="GI:34397424"
/translation="MSENKTYTEARRLIEIVRVIESPDVDELTKLAEALALGF
CREKLTADQKQIEELMAKLSWKNAML"
8123..8791
/gene="ispb"
/locus_tag="PGI434"
8123..8791
/gene="ispb"
/note="similar to GB:M96326, SP:P20160, PID:I79302, and
PID:28977; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="4-diphosphocytidylyl-2C-methyl-D-erythritol
synthase"
/protein_id="AA066487.1"
/db_xref="GI:34397425"
/translation="MEOKRYALIVAGGHGLRMGADRPKQFILLAGLPVLMHTLNRFP
HVDALVLPDTHAYWELCKRYDFSVHRVWAGGNTRFASVRNGLVVDPDGLVAV
HGVRLVSAETIDACPLAELKGAVPCRPMTESLRYATDGNVYVDRGRYVTVTP
QTRSEWLREATRPYEEYFTDDCSVYEHFGRPVALLVGNINIKULTPDLDSLAKL
LLTS"
9020..9096
/gene="tRNA-Asp-1"
9020..9096
/gene="tRNA-Asp-1"
/product="tRNA-Asp"
/note="codon recognized: GAC"
9404..10630
/locus_tag="PGI435"
9404..10630
/locus_tag="PGI435"
/codon_start=1
/transl_table=11

/product="integrase"
/protein_id="AA066488.1"
/db_xref="GI:34397426"
/translation="MKSQFRSVFYLRNINKEGKTPILIRIYNKRELISLGGTGLSV
NAQWLDSKEKVKGHSAAEVENKIEIRADILTIYKRLVETVDDITPERIKSEYCG
QDTLNSIVELFDKNEEDVRAQVINGINTAATLOKYENSKGHFTFLKAKNRTDLKFS
ELTELVINHFILYLLTVAHCCPNTATKLKLFKTIYGRHGFSLDHPFNKRPFHQ
DSSGFUDIEIQRIIENKIEISIRPLDLIVRDLIFSCSGLAYIDVSNLTEDNIVELDG
REWINTNRQKTRVATNIIILLIPKRIIEKVPVRKNGHLLPILSNQKNAYLKEIADL

Query Match 71.5%; Score 18.6; DB 1; Length 301618;
Best Local Similarity 84.0%; Pred. No. le+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGGAGGCCCGAATCTACGAA 26
||||| ||||| ||||| ||||| |||||
Db 48065 CAACGGCATGCCCGAATCAACGAA 48041

RESULT 30
BX569695 344615 bp DNA circular BCT 23-SEP-2003
LOCUS
Synecococcus sp. WH8102 complete genome; segment 7/7.
DEFINITION
BX569695 BX548020
ACCESSION
BX569695.1 GI:33639324
VERSION
complete genomes.
KEYWORDS
Synecococcus sp. WH 8102
SOURCE
Synecococcus sp. WH 8102
ORGANISM
Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
REFERENCE
1 (bases 1 to 344615)
AUTHORS
Palenik,B., Brahamsha,B., Larimer,F., Land,M., Hauser,L., Chain,P.,
Lamerdin,J., Regala,W., Allen,E.A., McCarren,J., Paulsen,I.,
Dufresne,A., Partensky,F., Webb,E. and Waterbury,J.
The genome of a motile marine Synecococcus
Nature 424 (6952), 1037-1042 (2003)
22825697
12917641
2 (bases 1 to 344615)
Larimer,F. and Palenik,B.
Synecococcus genome consortium
Direct Submission
Submitted (01-JUL-2003) Larimer, F., Submitted on behalf of the
Synecococcus genome consortium, the DOE Joint Genome Institute, CA
Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA
94598, USA, and the Genome Analysis Group, Oak Ridge National
Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov

FEATURES
source
1..344615 Synecococcus sp. WH 8102
/organism="Synecococcus sp. WH 8102"
/mol_type="genomic DNA"
/strain="WH8102"
/db_xref="taxon:84588"
67..1716
/gene="nadB"
/locus_tag="SYNW2192"
67..1716
/gene="nadB"
/locus_tag="SYNW2192"
/EC_number="1.4.3.16"
/note="putative assignment"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative L-aspartate oxidase"
/protein_id="CAE08707.1"
/db_xref="GI:33639325"

gene
CDS

```
IPVAPAAHYWGGVATDLOAATDPLGLYAVGEVACTVGHAYGANLNASLMLGLAEFAHR
LAETIDGPAIKPSNPAPPOAYGOALEGATISAGLMEQIEQLRLFCWQAGVDSRGERL
NAGLDRCELESLDOCLPLNCLORALLDDSSRDLANLLDLRLHLVTSKMLLEAC
LPQESRGCHFRDTRASSQWQCHSRQSRARGLHTRAVRF"
complement(1724..2853)
/locus_tag="SYNW2193"
complement(1724..2653)
/locus_tag="SYNW2193"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE08708.1"
/db_xref="GI:33639326"
/translations="MGTRLVSRRRDQAGAKWARIAMAVLATVGVIDTGSITLKRWGL
LGNLSCPMGADGCDVGLNSPWGTVDGIPLSFVGLYLAIVLMAILLPLGLAEKNS
DLASRTWGMFVSLGWVPSGLLGMFKIOAFCEFCVLSLFLSLVLSLIGGG
WDPPGGLIFRGLLALAVLLGLLWASVDPDEPATAFGPVAVPVVTTASSTAILAL
ADHLTASGVAIYACWPCHEQEQLFGKATKQLKVVCEAGDRNNQDLCKRSKLE
GFSWEINGELSDSGVLPLEVLAJLSGYQGDRE"
order(complement(2372..2437),complement(2261..2326),
complement(2183..2239),complement(2087..2152))
/locus_tag="SYNW2193"
/notes="4 probable transmembrane helices predicted by
TMHMM2.0"
2721..4115
/locus_tag="SYNW2194"
2721..4115
/locus_tag="SYNW2194"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="possible Fe-S oxidoreductase"
/protein_id="CAE08709.1"
/db_xref="GI:33639327"
/translations="MAGGAVMTPTPTKPTVAPAHLCENKRVDTTEHMGVLLAQAGVY
STNENDAVVVVNTCSFIQDARESVRTLVAEQKELIAGCLAQHFQDSLESIP
EAKYVGTGDOYHIVLQVGEVGERVNRVSAPTVFGDEHLPQRSTTTQAVAFKVA
EGCDYACAFICIPALGQKRSRETSIAVAHQVQGVQGLILISQITINYGLLDLY
KPKLAELRALGEIVFWRVHYAYPTGLTPEVLAARYBNVVPYLDLPLOHSHPEV
LRANRPQWADNRLLDQIDQDPVALRTLIVFGPGETEEHFQHLMEFLRRQRFD
HVGVTFSFSDGTAAELPNPDPVDAQKDALMALQOPISAERNHSMVSVRTDVLI
EOHPOTGOMIGSCARFAPEVDGEVHQPCDQKAAPTGTVMPVQITGADVLDLSRI
VGARDMVAATRAIDA"
4112..5341
/locus_tag="SYNW2195"
4112..5341
/locus_tag="SYNW2195"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative multidrug efflux MFS transporter"
/protein_id="CAE08710.1"
/db_xref="GI:33639328"
/translations="NRAFLRLDQQRSLFVWSIGTAGSFAGITAKGWILLHGSADP
FVLNFAALSPSLISGPAGVTRDLCGETVLVQAWGLLAAALGALSIPLEGT
AOLLILLGSTVIGAGAPETARNKVCILLVEEPKOLAGYLTFSVFNVGKLYGPP
IGGLWAMAGPAGALGIDAAVYLLPIASVYMLLOPERDMEQSRGDDATLLKAWDC
GSTLRGLVILTAVLCVNFPHGLAPLJADQVLGPDPRDLGFTSVLAGSGIAGVIL
QRCSTRSRPFLLGCPGLITAIACQGMAGSSVAISUSMAFAGTAGLISCNL
ITQIGAPQWNRGMAGLSQIAFIGGGGIGSLAALDVMTATSATFALTGGIGAVIAL
LWIRRGGVLEPRLSA"
4112..4198
/locus_tag="SYNW2195"
/notes="Signal predicted by SignalP 2.0 HMM (signal peptide
probability 0.967) with cleavage site probability 0.931 at
residue 23"
misc_feature
order(4148..4216,4244..4312,4331..4399,4412..4480,
4541..4609,4622..4690,4772..4840,4883..4939,4973..5032,
5060..5119,5156..5224,5237..5296)
/locus_tag="SYNW2195"
/notes="12 probable transmembrane helices predicted by
TMHMM2.0"

5566..7140
/locus_tag="SYNW2196"
5566..7140
/locus_tag="SYNW2196"
/EC_number="2.1.1.-"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE08711.1"
/db_xref="GI:33639329"
/translations="MRTLFIYFEPFKTWSYEKILELVNRKVLPLPLGMVTVAALLPQ
DMEMKLVDRNVDRDVTTEENWAEWIIISGMIVQDDNAVOISKAKORGLPVAIGGPPA
SSTPAPLDOADPKILDEGITLPMFLAALERGDTSGRTAEGDKPVTVPPIPRFD
LLELDAYDSMSVQSRGCPFCNCFCDIIVLXGRKPTKTPSQVLAELQSLYDLGWRGS
LFLVNDNFGKRNKAKLLPQIRTWQBERGYFPFATASVDLADDERMMQMDHARE
RSVFLGCTPRESSLETARKICNTRNPLDAAVDRTANGIRVMAGFIIGFQEKDGAG
HRIVDFVTRTGIPAMNMQLQALPRTALWASLEKEGRLIQGEDAKGNQNLNLFKP
TRPDIANDYVDPACALYEFNAYMDRVYSLKMGAPRWAKAALPTMTDTRALSIV
IWRQIKRDTSRFRWYKLLGMARONPALLEQFLVVLAHNEHFEYRSIVQOEIRQLE
SLPPEPSAQXELQPV"
complement(7144..7551)
/locus_tag="SYNW2197"
complement(7144..7551)
/locus_tag="SYNW2197"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE08712.1"
/db_xref="GI:33639330"
/translations="MVKTPPEPASLTVAIDALDERLSQRTALDPSGVFLIKVADAG
ELVLEHYGNTIDEXGLARDSDTGEVLSCKGGNGLRQPSRIYRGSSAKQVGLTEGDC
PHPLSRDLHALYLQRELQAKAQCLINGSAYVQD"
complement(7561..8736)
/locus_tag="SYNW2198"
complement(7561..8736)
/locus_tag="SYNW2198"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE08713.1"
/db_xref="GI:33639331"
/translations="MSSLTARHRSIQEIPQHQWALVSDHAIFPYRWSWLEAESG
SIIPEQWQPLHLALWRGDEIVAPLYLKGHSYGEFVDQTFARLADGLGRYPKL
LGMSPFVPLGYRFRHVLPGEDAEALTVLLAIDAFCEHNSILSCNFLYDQWRPLA
EAGACATWLNQOQSLWSRGDDQSDFDYLNFNANORRNKRKAKAVAGLTVTLTGE
OLDLELFOIMHGFYEQHCARQWPGSKYLEQGFEEALAGLHREOVVLFSAHRGDPREP
VAMSMCVDDHHLNRYWGSSEEDCLCHFVCYVAPLEWASIQSIDSPDPCGAGSHKR
RRGVARPHASLHRWYQPMQDLIRAMLPKYNGLMLEIEINALELFPKADAPSLSL"
complement(8793..9095)
/locus_tag="SYNW2199"
complement(8793..9095)
/locus_tag="SYNW2199"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical"
/protein_id="CAE08714.1"
/db_xref="GI:33639332"
/translations="MAGDVIAITGCKYRVWGGEDNDTFTKLDGCGKGMKIMDPEAGD
SITFCGCASTRIEQRGNKNAWIKNDVKAIVYGVTAADLQIDFDQAIITWADFLA"
complement(9250..9931)
/locus_tag="SYNW2200"
complement(9250..9931)
/locus_tag="SYNW2200"
/EC_number="1.1.1.193"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative riboflavin-specific deaminase"
/protein_id="CAE08715.1"
```


GenCore version 5.1.6

Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 149.397 Seconds
(without alignments)

739.327 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggcagggccgaattctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseqn_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26	100.0	393	2	AAT49174
C 2	26	100.0	110000	4	AAI99682_42
C 3	26	100.0	110000	4	AAI99683_42
C 4	18.6	71.5	2623	4	ABL11529
C 5	18.6	71.5	6747	4	ABL11528
C 6	18.6	71.5	14705	4	AAI99523
C 7	18.6	71.5	14705	7	ACF64452
C 8	18.2	70.0	546	6	ABQ90865
C 9	18	69.2	656	6	ABQ90865 M. capsul
C 10	18	69.2	656	6	ABQ18285
C 11	18	69.2	1113	6	ABQ18284
C 12	18	69.2	1113	6	ABQ46934
C 13	18	69.2	2050	6	ABQ46935
C 14	18	69.2	2050	6	ABQ39756
C 15	18	69.2	110000	4	AAI99682_04
C 16	18	69.2	110000	4	AAI99682_05
C 17	18	69.2	110000	4	AAI99683_04
C 18	18	69.2	110000	4	AAI99683_05
C 19	17.6	67.7	839	6	ABQ16680
C 20	17.6	67.7	839	6	ABQ16681
C 21	17.6	67.7	851	6	ABQ33932
C 22	17.6	67.7	851	6	ABQ33933
C 23	17.6	67.7	1017	7	ACA24162

24	17.6	67.7	11340	4	ABU16852
C 25	17.6	67.7	110000	2	AAV30458_0
C 26	17.6	67.7	110000	2	AAV30459_0
C 27	17.2	66.2	349980	6	ABQ18444
C 28	17	65.4	530	6	ABQ14535
C 29	17	65.4	530	6	ABQ14534
C 30	17	65.4	879	3	AAAS5205
C 31	17	65.4	913	3	AAAF08967
C 32	17	65.4	934	6	ABQ45717
C 33	17	65.4	934	6	ABQ45716
C 34	17	65.4	1173	7	ACA25903
C 35	17	65.4	1225	4	ABL08271
C 36	17	65.4	4568	4	ABL08270
C 37	17	65.4	4432	3	AAAS5187
C 38	17	65.4	93483	8	ADA03083
C 39	17	65.4	93483	8	ADA66367
C 40	17	65.4	93483	9	ADB72821
C 41	17	65.4	349980	6	ABQ18447
C 42	16.8	64.6	637	7	A3253960
C 43	16.8	64.6	1230	1	AAAF08967
C 44	16.8	64.6	1230	1	AAAF08967
C 45	16.8	64.6	1230	1	AAAF08967
C 46	16.8	64.6	1755	7	ADA70624
C 47	16.8	64.6	5274	1	AAAF0679
C 48	16.8	64.6	5274	1	AAAF0824
C 49	16.6	63.8	290	6	ABL73770
C 50	16.6	63.8	932	6	ABQ34921
C 51	16.6	63.8	932	6	ABQ34920
C 52	16.6	63.8	1431	4	ABL08403
C 53	16.6	63.8	1455	7	ACA30126
C 54	16.6	63.8	2031	9	ADD46116
C 55	16.6	63.8	2031	9	ABE61715
C 56	16.6	63.8	2157	7	ACA27237
C 57	16.6	63.8	2307	4	ABL23687
C 58	16.6	63.8	3077	7	ABZ10070
C 59	16.6	63.8	4218	9	ADG87164
C 60	16.6	63.8	5479	4	ABL23686
C 61	16.6	63.8	6077	6	ABL33855
C 62	16.6	63.8	6077	9	ADB54168
C 63	16.6	63.8	6077	9	ADB84126
C 64	16.6	63.8	6077	9	ADB84126
C 65	16.6	63.8	7543	8	AAAL61505
C 66	16.6	63.8	7543	3	AAAS3978
C 67	16.6	63.8	34769	4	AAAS46775
C 68	16.6	63.8	67832	8	ADA02801
C 69	16.6	63.8	67832	9	ADB72539
C 70	16.6	63.8	67832	9	ADC85281
C 71	16.6	63.8	156638	6	ABQ81850
C 72	16.4	63.1	511	6	ABQ50581
C 73	16.4	63.1	511	6	ABQ50580
C 74	16.4	63.1	522	6	ABQ33181
C 75	16.4	63.1	522	6	ABQ33180
C 76	16.4	63.1	566	6	ABQ50784
C 77	16.4	63.1	566	6	ABQ50785
C 78	16.4	63.1	595	3	AAAC38541
C 79	16.4	63.1	613	6	ABQ29765
C 80	16.4	63.1	613	6	ABQ29764
C 81	16.4	63.1	643	6	ABQ28009
C 82	16.4	63.1	643	6	ABQ28008
C 83	16.4	63.1	699	7	ACA42767
C 84	16.4	63.1	717	6	ABQ23027
C 85	16.4	63.1	717	6	ABQ23026
C 86	16.4	63.1	723	6	ABQ40140
C 87	16.4	63.1	723	6	ABQ40141
C 88	16.4	63.1	734	6	ABQ38865
C 89	16.4	63.1	734	6	ABQ38864
C 90	16.4	63.1	748	6	ABQ24376
C 91	16.4	63.1	748	6	ABQ24377
C 92	16.4	63.1	798	6	ABQ29802
C 93	16.4	63.1	798	6	ABQ29803
C 94	16.4	63.1	836	6	ABQ13596
C 95	16.4	63.1	836	6	ABQ13597
C 96	16.4	63.1	846	6	ABQ46632

AbL16852	Drosophil
AAV30458	Rhizobium
AAV30459	Rhizobium
ABQ18444	Bifidobac
ABQ14535	Oligonuc
ABQ14534	Oligonuc
AAAS5205	C. symbio
AAAF08967	Fusarium
ABQ45717	Oligonuc
ABQ45716	Oligonuc
ACA25903	Prokaryot
ABL08271	Drosophil
ABL08270	Drosophil
AAAS5187	Cenarchae
ADA03083	Mouse mCG
ADA66367	Mouse mCG
ADB72821	Mouse mCG
ABQ18447	Bifidobac
ABZ53960	Aspergill
AAAF08967	Sequence
AAAF08967	Sequence
ADA70624	Rice gene
AAAF0679	Sequence
AAAF0824	Catechol
ABL73770	Corn tass
ABQ34921	Oligonuc
ABQ34920	Oligonuc
ABL08403	Drosophil
ACA30126	Prokaryot
ADD46116	Rat gene
ABE61715	Rat gene
ACA27237	Prokaryot
ABZ10070	Drosophil
ADG87164	Haematopo
ABL23686	Human GFC
ABL33855	Drosophil
ADB54168	Human imm
ADB84126	Human imm
AAAL61505	Human lym
AAAS3978	Drosophil
AAAS46775	Drosophil
ADA02801	Mouse itk
ADB72539	Mouse itk
ADC85281	Mouse itk
ABQ81850	Bifidobac
ABQ50581	Oligonuc
ABQ50580	Oligonuc
ABQ33181	Oligonuc
ABQ33180	Oligonuc
ABQ50784	Oligonuc
ABQ50785	Oligonuc
AAAC38541	Arabidops
ABQ29765	Oligonuc
ABQ29764	Oligonuc
ABQ28009	Oligonuc
ABQ28008	Oligonuc
ACA42767	Prokaryot
ABQ23027	Oligonuc
ABQ23026	Oligonuc
ABQ40140	Oligonuc
ABQ40141	Oligonuc
ABQ38865	Oligonuc
ABQ38864	Oligonuc
ABQ24376	Oligonuc
ABQ24377	Oligonuc
ABQ29802	Oligonuc
ABQ29803	Oligonuc
ABQ13596	Oligonuc
ABQ13597	Oligonuc
ABQ46632	Oligonuc

97	16.4	63.1	846	6	ABQ46633	Oligonuc1	170	16	61.5	754	6	ABQ34323	Oligonuc1
98	16.4	63.1	874	6	ABQ44403	Oligonuc1	171	16	61.5	782	6	ABQ53104	Oligonuc1
99	16.4	63.1	876	6	ABQ44402	Oligonuc1	172	16	61.5	782	6	ABQ53105	Oligonuc1
100	16.4	63.1	877	6	ABQ29264	Oligonuc1	173	16	61.5	897	7	ADA70243	Rice gene
101	16.4	63.1	877	6	ABQ29265	Oligonuc1	174	16	61.5	897	8	ADA48201	Rice gene
102	16.4	63.1	976	6	ABQ52966	Oligonuc1	175	16	61.5	941	3	AC95102	Cat flea
103	16.4	63.1	976	6	ABQ52967	Oligonuc1	176	16	61.5	969	7	ACA25173	Prokaryot
104	16.4	63.1	976	6	ABQ52967	Oligonuc1	177	16	61.5	969	7	ACA25173	Prokaryot
105	16.4	63.1	998	6	ABQ33244	Oligonuc1	178	16	61.5	971	2	AAT39801	Human clo
106	16.4	63.1	998	6	ABQ33245	Oligonuc1	179	16	61.5	984	6	ABQ27350	Oligonuc1
107	16.4	63.1	1204	6	ABQ14447	Oligonuc1	180	16	61.5	984	6	ABQ27351	Oligonuc1
108	16.4	63.1	1204	6	ABQ14446	Oligonuc1	181	16	61.5	984	9	ADD82096	Human ama
109	16.4	63.1	1240	6	ABQ42114	Oligonuc1	182	16	61.5	1080	7	ACA43613	Prokaryot
110	16.4	63.1	1240	6	ABQ42115	Oligonuc1	183	16	61.5	1227	6	ABK73883	Prokaryot
111	16.4	63.1	1312	6	ABQ30092	Oligonuc1	184	16	61.5	1242	4	AA54493	Botulinum
112	16.4	63.1	1312	6	ABQ30093	Oligonuc1	185	16	61.5	1457	4	AAQ92642	Human Nck
113	16.4	63.1	1422	7	ACA38029	Prokaryot	186	16	61.5	1737	6	ABA05826	Arthrobac
114	16.4	63.1	1579	6	ABQ35414	Oligonuc1	187	16	61.5	1737	6	ABA05825	Arthrobac
115	16.4	63.1	1681	6	ABQ35415	Oligonuc1	188	16	61.5	1761	4	ABL18843	Drosophil
116	16.4	63.1	1681	6	ABQ33738	Oligonuc1	189	16	61.5	1884	6	ABA05824	Arthrobac
117	16.4	63.1	1726	5	AA572093	Oligonuc1	190	16	61.5	1956	5	AA505824	Arthrobac
118	16.4	63.1	2000	6	ABZ17079	Arabidops	191	16	61.5	1985	2	AA846777	DNA encod
119	16.4	63.1	2265	7	AAH65400	C glutami	192	16	61.5	2157	6	ABQ30957	Fragment
120	16.4	63.1	2265	7	ACA02105	C glutami	193	16	61.5	2157	6	ABQ30957	Oligonuc1
121	16.4	63.1	2670	7	ABT17795	Aspergill	194	16	61.5	2157	6	ABQ30956	Oligonuc1
122	16.4	63.1	2837	7	ABT17795	Aspergill	195	16	61.5	2267	2	AA84665	Human KDR
123	16.4	63.1	3499	5	AA580109	Human nov	196	16	61.5	2364	7	ACA23492	Prokaryot
124	16.4	63.1	3499	5	ADC32374	Human nov	197	16	61.5	2559	6	ABK98547	Botulinum
125	16.4	63.1	3969	5	AA580885	DNA encod	198	16	61.5	2918	3	AAZ44322	S. tendae
126	16.4	63.1	3969	5	AA580885	DNA encod	199	16	61.5	3300	4	AAH42258	Streptomy
127	16.4	63.1	6249	6	ABK31305	Signal tr	200	16	61.5	3318	7	ACA43477	Nucleotid
128	16.4	63.1	6249	6	ABK31305	Signal tr	201	16	61.5	3657	7	ADA69347	Rice gene
129	16.4	63.1	6249	6	ABK31305	Signal tr	202	16	61.5	3658	4	ABL14055	Prokaryot
130	16.4	63.1	6249	6	ABK31305	Signal tr	203	16	61.5	4079	7	ACA56650	Signalin
131	16.4	63.1	6652	6	ABK66987	Human ang	204	16	61.5	4133	4	ABL18942	Drosophil
132	16.4	63.1	7048	4	ABK66987	Human ang	205	16	61.5	4519	9	ABK84127	Human cdn
133	16.4	63.1	7048	4	ABK66987	Human ang	206	16	61.5	4519	9	ABK84127	Human cdn
134	16.4	63.1	7967	7	ABK66987	Human ang	207	16	61.5	4851	7	ACA37619	Prokaryot
135	16.4	63.1	8300	5	AA661281	N. magada	208	16	61.5	5123	4	ABL21202	Drosophil
136	16.4	63.1	8300	5	AA661281	N. magada	209	16	61.5	5123	4	ABL21202	Drosophil
137	16.4	63.1	349980	5	AAH68525	C glutami	210	16	61.5	12008	4	ABL21202	Drosophil
138	16.4	63.1	349980	5	AAH68525	C glutami	211	16	61.5	12008	4	ABL21202	Drosophil
139	16.2	62.3	330	6	ABQ81846	Bifidobac	212	16	61.5	15377	2	AAQ25375	Prokaryot
140	16.2	62.3	471	5	AAH67419	C glutami	213	16	61.5	15377	2	AAQ25375	Prokaryot
141	16.2	62.3	471	5	AAH67419	C glutami	214	16	61.5	15377	2	AAQ25375	Prokaryot
142	16.2	62.3	471	5	AAH67419	C glutami	215	16	61.5	15377	2	AAQ25375	Prokaryot
143	16.2	62.3	471	5	AAH67419	C glutami	216	16	61.5	15377	2	AAQ25375	Prokaryot
144	16.2	62.3	471	5	AAH67419	C glutami	217	16	61.5	15377	2	AAQ25375	Prokaryot
145	16.2	62.3	471	5	AAH67419	C glutami	218	16	61.5	15377	2	AAQ25375	Prokaryot
146	16.2	62.3	471	5	AAH67419	C glutami	219	16	61.5	15377	2	AAQ25375	Prokaryot
147	16.2	62.3	471	5	AAH67419	C glutami	220	16	61.5	15377	2	AAQ25375	Prokaryot
148	16.2	62.3	471	5	AAH67419	C glutami	221	16	61.5	15377	2	AAQ25375	Prokaryot
149	16.2	62.3	471	5	AAH67419	C glutami	222	16	61.5	15377	2	AAQ25375	Prokaryot
150	16.2	62.3	471	5	AAH67419	C glutami	223	16	61.5	15377	2	AAQ25375	Prokaryot
151	16.2	62.3	471	5	AAH67419	C glutami	224	16	61.5	15377	2	AAQ25375	Prokaryot
152	16.2	62.3	471	5	AAH67419	C glutami	225	16	61.5	15377	2	AAQ25375	Prokaryot
153	16.2	62.3	471	5	AAH67419	C glutami	226	16	61.5	15377	2	AAQ25375	Prokaryot
154	16.2	62.3	471	5	AAH67419	C glutami	227	16	61.5	15377	2	AAQ25375	Prokaryot
155	16.2	62.3	471	5	AAH67419	C glutami	228	16	61.5	15377	2	AAQ25375	Prokaryot
156	16.2	62.3	471	5	AAH67419	C glutami	229	16	61.5	15377	2	AAQ25375	Prokaryot
157	16.2	62.3	471	5	AAH67419	C glutami	230	16	61.5	15377	2	AAQ25375	Prokaryot
158	16.2	62.3	471	5	AAH67419	C glutami	231	16	61.5	15377	2	AAQ25375	Prokaryot
159	16.2	62.3	471	5	AAH67419	C glutami	232	16	61.5	15377	2	AAQ25375	Prokaryot
160	16.2	62.3	471	5	AAH67419	C glutami	233	16	61.5	15377	2	AAQ25375	Prokaryot
161	16.2	62.3	471	5	AAH67419	C glutami	234	16	61.5	15377	2	AAQ25375	Prokaryot
162	16.2	62.3	471	5	AAH67419	C glutami	235	16	61.5	15377	2	AAQ25375	Prokaryot
163	16.2	62.3	471	5	AAH67419	C glutami	236	16	61.5	15377	2	AAQ25375	Prokaryot
164	16.2	62.3	471	5	AAH67419	C glutami	237	16	61.5	15377	2	AAQ25375	Prokaryot
165	16.2	62.3	471	5	AAH67419	C glutami	238	16	61.5	15377	2	AAQ25375	Prokaryot
166	16.2	62.3	471	5	AAH67419	C glutami	239	16	61.5	15377	2	AAQ25375	Prokaryot
167	16.2	62.3	471	5	AAH67419	C glutami	240	16	61.5	15377	2	AAQ25375	Prokaryot
168	16.2	62.3	471	5	AAH67419	C glutami	241	16	61.5	15377	2	AAQ25375	Prokaryot
169	16.2	62.3	471	5	AAH67419	C glutami	242	16	61.5	15377	2	AAQ25375	Prokaryot


```
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACACGGCGAGCGCCGAATCTACGAA 26
Db 26924 ACACGGCGAGCGCCGAATCTACGAA 26949

RESULT 3
AAI99683_42
Continuation (43 of 44) of AAI99683 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACACGGCGAGCGCCGAATCTACGAA 26
Db 19170 ACACGGCGAGCGCCGAATCTACGAA 19195

RESULT 4
ABL11529
ID ABL11529 standard; cDNA; 2623 BP.
XX
AC ABL11529;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29069.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEXE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB67426.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 29069; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
```

XX SQ Sequence 2623 BP; 741 A; 673 C; 694 G; 515 T; 0 U; 0 Other;
Query Match 71.5%; Score 18.6; DB 4; Length 2623;
Best Local Similarity 84.0%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGGAATCTACGA 25
DB 1566 AGACGGCGAGCGCCGGAATCTCGA 1590
RESULT 5
ABL11528/c
ID ABL11528 standard; cDNA; 6747 BP.
XX AC ABL11528;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29066.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
WPI; 2001-656860/75.
P-PSDB; ABB67425.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
Claim 1; SEQ ID NO 29066; 21pp + Sequence Listing; English.
The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 6747 BP; 1672 A; 1591 C; 1660 G; 1824 T; 0 U; 0 Other;
Query Match 71.5%; Score 18.6; DB 4; Length 6747;
Best Local Similarity 84.0%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGGAATCTACGA 25
DB 2375 AGACGGCGAGCGCCGGAATCTCGA 2351
RESULT 6
AAS59523

ID AAS59523 standard; DNA; 14705 BP.
XX AAS59523;
XX DT 13-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein encoding DNA #18.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
XX OS Propionibacterium acnes.
XX WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US012865.
XX PR 21-APR-2000; 2000US-0199047P.
XX PR 02-JUN-2000; 2000US-0208841P.
XX PR 07-JUL-2000; 2000US-0216747P.
XX PA (CORI-) CORIXA CORP.
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
WPI; 2001-616774/71.
XX Propionibacterium acnes polypeptides and nucleic acids useful for
vaccinating against and diagnosing infections, especially useful for
treating acne vulgaris.
Claim 1; SEQ ID NO 18; 1069pp; English.
Sequences AAS59506-AAS59804 represent DNA molecules encoding
Propionibacterium acnes immunogenic polypeptides. The proteins and their
associated DNA sequences are used in the treatment, prevention and
diagnosis of medical conditions caused by P. acnes. The disorders include
SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
infections of bone, joints and the central nervous system, however it is
particularly involved in the inflammatory lesions associated with acne
vulgaris. A method for detecting the presence or absence of P. acnes in a
patient comprises contacting a sample with a binding agent that binds to
the proteins of the invention and determining the amount of bound protein
in the sample. The polypeptides may be used as antigens in the production
of antibodies specific for P. acnes proteins. These antibodies can be
used to downregulate expression and activity of P. acnes polypeptides and
therefore treat P. acnes infections. The antibodies may also be used as
diagnostic agents for determining P. acnes presence, for example, by
enzyme linked immunosorbent assay (ELISA). This sequence encodes the
polypeptides shown in AAU44569-AAU44720 and AAU67491-AAU67493. Note: The
sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;
Query Match 71.5%; Score 18.6; DB 4; Length 14705;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGGAATCTACGA 25
DB 4255 ACAACGGCGAGCGCCGGAATCTACGA 4279
RESULT 7
ACF64452

ACF64452 standard; DNA; 14705 BP.
ACF64452;
17-OCT-2003 (first entry)
Propionibacterium acnes DNA contig sequence #18.
Acne vulgaris; antisephorrhoeic; dermatological; antibacterial;
immunostimulant; immune response; vaccine; ds.
Propionibacterium acnes.
WO2003033515-A1.
24-APR-2003.
11-OCT-2002; 2002WO-US032727.
15-OCT-2001; 2001US-00978825.
(CORI-) CORIXA CORP.
Mitchem JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
Barth B, Vallie-Douglass J;
WPI; 2003-381789/36.
New Propionibacterium acnes polypeptides and polynucleotides encoding the
or for diagnosing, preventing or treating acne vulgaris,
or for stimulating an immune response specific for a P. acnes protein.
Claim 1; SEQ ID NO 18; 1481pp; English.
The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
encoding a Propionibacterium acnes protein. The invention also relates to
polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
immunogenic fragments of P. acnes polypeptides. The invention
additionally encompasses expression vectors and host cells comprising a
polynucleotide of the invention; antibodies against polypeptides of the
invention; fusion proteins comprising a polypeptide of the invention; a
method for stimulating an immune response specific for a P. acnes
polypeptide and an isolated T cell population comprising T cells prepared
via this method; a vaccine composition (comprising P. acnes polypeptides,
polynucleotides, antibodies, fusion proteins, T cell populations, or
antigen-presenting cells that express the polypeptide); a method and kit
for detecting or determining the presence or absence of P. acnes in a
patient; and a method for inhibiting the development of P. acnes in a
patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
proteins, T cell populations or antigen-presenting cells that express the
polypeptides are useful for diagnosing, preventing or treating acne
vulgaris, or for stimulating an immune response specific for a P. acnes
protein. The polynucleotides can also be used as probes or primers for
nucleic acid hybridisation. The vaccine composition is useful for the
stimulation of an immune response against P. acnes, or for treating acne,
and the kit is useful for performing a diagnostic assay. The present
sequence represents a P. acnes DNA contig which is specifically claimed
in the invention. Note: The sequence data for this patent did not form
part of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;
Query Match 71.5%; Score 18.6; DB 7; Length 14705;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db 4255 ACAACGTCGTCGCCCGCAATCTACGA 4279
ACF64452 standard; DNA; 14705 BP.
ACF64452;
17-OCT-2003 (first entry)
Propionibacterium acnes DNA contig sequence #18.
Acne vulgaris; antisephorrhoeic; dermatological; antibacterial;
immunostimulant; immune response; vaccine; ds.
Propionibacterium acnes.
WO2003033515-A1.
24-APR-2003.
11-OCT-2002; 2002WO-US032727.
15-OCT-2001; 2001US-00978825.
(CORI-) CORIXA CORP.
Mitchem JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
Barth B, Vallie-Douglass J;
WPI; 2003-381789/36.
New Propionibacterium acnes polypeptides and polynucleotides encoding the
or for diagnosing, preventing or treating acne vulgaris,
or for stimulating an immune response specific for a P. acnes protein.
Claim 1; SEQ ID NO 18; 1481pp; English.
The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
encoding a Propionibacterium acnes protein. The invention also relates to
polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
immunogenic fragments of P. acnes polypeptides. The invention
additionally encompasses expression vectors and host cells comprising a
polynucleotide of the invention; antibodies against polypeptides of the
invention; fusion proteins comprising a polypeptide of the invention; a
method for stimulating an immune response specific for a P. acnes
polypeptide and an isolated T cell population comprising T cells prepared
via this method; a vaccine composition (comprising P. acnes polypeptides,
polynucleotides, antibodies, fusion proteins, T cell populations, or
antigen-presenting cells that express the polypeptide); a method and kit
for detecting or determining the presence or absence of P. acnes in a
patient; and a method for inhibiting the development of P. acnes in a
patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
proteins, T cell populations or antigen-presenting cells that express the
polypeptides are useful for diagnosing, preventing or treating acne
vulgaris, or for stimulating an immune response specific for a P. acnes
protein. The polynucleotides can also be used as probes or primers for
nucleic acid hybridisation. The vaccine composition is useful for the
stimulation of an immune response against P. acnes, or for treating acne,
and the kit is useful for performing a diagnostic assay. The present
sequence represents a P. acnes DNA contig which is specifically claimed
in the invention. Note: The sequence data for this patent did not form
part of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;
Query Match 71.5%; Score 18.6; DB 7; Length 14705;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db 4255 ACAACGTCGTCGCCCGCAATCTACGA 4279

RESULT 8
ABQ90865
ID ABQ90865 standard; DNA; 546 BP.
XX
AC ABQ90865;
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #850 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
PN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002WO-NO000019.
XX
PR 12-JAN-2001; 2001NO-00000235.
XX
PR 12-JAN-2001; 2001NO-00000239.
XX
PA (UNIF-) UNIFOB STIFTTELSEN UNIV BERGEN.
XX
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HS, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
WPI; 2002-557818/59.
XX
Novel DNA array useful for determining differential expression of
Methylococcus capsulatus genes, comprises polynucleotides or
oligonucleotides representative for a selective number of Methylococcus
capsulatus genes.
XX
Claim 14; Page 374; 678pp; English.
XX
The invention relates to a novel DNA array giving a representation of a
number of Methylococcus capsulatus genes. The method of the invention is
useful for determination of the differential expression of the genes of
M. capsulatus, and for studying gene expression on a genomic scale and in
gene expression assays of M. capsulatus genes. The sequences shown in
ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
invention
XX
Sequence 546 BP; 100 A; 208 C; 164 G; 74 T; 0 U; 0 Other;
Query Match 70.0%; Score 18.2; DB 6; Length 546;
Best Local Similarity 87.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CAACGGCGAGGCCCGCAATCTACG 24
Db 126 CAACGGCGAGGCCCGCAATCTACG 148

RESULT 9
ABQ18285
ID ABQ18285 standard; DNA; 656 BP.
XX
AC ABQ18285;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4876.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX

PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC status of many C residues to be determined simultaneously. ABQ13410-
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1113 BP; 162 A; 140 C; 418 G; 393 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 1113;
Best Local Similarity 80.8%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26
Db 765 AAAACGACGACGACCGAACTACGAA 740
RESULT 12
ABQ46935
ID ABQ46935 standard; DNA; 1113 BP.
AC ABQ46935;
XX
XX
DT 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 33526.
DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
XX
XX
PD 07-MAR-2002.
XX
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
XX
PA (EPIG-) EPIGENOMICS AG.

PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC status of many C residues to be determined simultaneously. ABQ13410-
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1113 BP; 393 A; 418 C; 140 G; 162 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 1113;
Best Local Similarity 80.8%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26
Db 349 AAAACGACGACGACCGAACTACGAA 374
RESULT 13
ABQ39756/c
ID ABQ39756 standard; DNA; 2050 BP.
AC ABQ39756;
XX
XX
DT 12-JUL-2002 (first entry)
XX
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26347.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
XX
XX
PD 07-MAR-2002.
XX
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX XX
DR WPI; 2002-371829/40.
XX XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX XX
SQ Sequence 2050 BP; 241 A; 276 C; 766 G; 767 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 2050;
Best Local Similarity 80.8%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGCGCCGAACTACGAA 26
Db 1203 ATAAACGGCGAGCGCCGAACTACGAA 1178
RESULT 14
ABQ39757
ID ABQ39757 standard; DNA; 2050 BP.
XX AC ABQ39757;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26348.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX FN WO200218632-A2.
XX PD 07-MAR-2002.
XX XX 01-SEP-2001; 2001WO-EP010074.
XX XX 01-SEP-2000; 2000DE-01043826.
XX PR 05-SEP-2000; 2000DE-01044543.
XX XX (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX XX

DR WPI; 2002-371829/40.
XX XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX XX
SQ Sequence 2050 BP; 767 A; 766 C; 276 G; 241 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 2050;
Best Local Similarity 80.8%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGCGCCGAACTACGAA 26
Db 848 ATAAACGGCGAGCGCCGAACTACGAA 873
RESULT 15
AAI99682_04
Continuation (5 of 45) of AAI99682 from base 400001 (Mycobacterium tuberculosis strain H
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 113000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000

RESULT 18
 AAI99683_05
 Continuation (6 of 44) of AAI99683 from base 500001 (Mycobacterium tuberculosis strain H
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
 WP Fragment Name Begin End
 WP AAI99683_00 110000 210000
 WP AAI99683_01 100001 210000
 WP AAI99683_02 200001 310000
 WP AAI99683_03 300001 410000
 WP AAI99683_04 400001 510000
 WP AAI99683_05 500001 610000
 WP AAI99683_06 600001 710000
 WP AAI99683_07 700001 810000
 WP AAI99683_08 800001 910000
 WP AAI99683_09 900001 1010000
 WP AAI99683_10 1000001 1110000
 WP AAI99683_11 1100001 1210000
 WP AAI99683_12 1200001 1310000
 WP AAI99683_13 1300001 1410000
 WP AAI99683_14 1400001 1510000
 WP AAI99683_15 1500001 1610000
 WP AAI99683_16 1600001 1710000
 WP AAI99683_17 1700001 1810000
 WP AAI99683_18 1800001 1910000
 WP AAI99683_19 1900001 2010000
 WP AAI99683_20 2000001 2110000
 WP AAI99683_21 2100001 2210000
 WP AAI99683_22 2200001 2310000
 WP AAI99683_23 2300001 2410000
 WP AAI99683_24 2400001 2510000
 WP AAI99683_25 2500001 2610000
 WP AAI99683_26 2600001 2710000
 WP AAI99683_27 2700001 2810000
 WP AAI99683_28 2800001 2910000
 WP AAI99683_29 2900001 3010000
 WP AAI99683_30 3000001 3110000
 WP AAI99683_31 3100001 3210000
 WP AAI99683_32 3200001 3310000
 WP AAI99683_33 3300001 3410000
 WP AAI99683_34 3400001 3510000
 WP AAI99683_35 3500001 3610000
 WP AAI99683_36 3600001 3710000
 WP AAI99683_37 3700001 3810000
 WP AAI99683_38 3800001 3910000
 WP AAI99683_39 3900001 4010000
 WP AAI99683_40 4000001 4110000
 WP AAI99683_41 4100001 4210000
 WP AAI99683_42 4200001 4310000
 WP AAI99683_43 4300001 4403765

Query Match 69.2%; Score 18; DB 4; Length 110000;
 Best Local Similarity 80.8%; Pred. No. 1.4e+02;
 Matches 2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGAA 26
 DB 7446 AGACCGGCGAGGTGCGAATCAACGAA 7471

RESULT 19
 ABQ16680/c
 ID ABQ16680 standard; DNA; 839 BP.
 XX AC
 XX ABQ16680;
 XX
 DT 12-JUL-2002 (first entry)
 DE
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 3271.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX PA (EPIT-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX DR
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 839 BP; 96 A; 89 C; 284 G; 370 T; 0 U; 0 Other;
 Query Match 67.7%; Score 17.6; DB 6; Length 839;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGCGCCGAATCTACGAA 26
 DB 458 AACGACGACGCGCGAATCTAAAAA 435

RESULT 20
 ABQ16681
 ID ABQ16681 standard; DNA; 839 BP.
 XX AC
 XX ABQ16681;
 XX
 DT 12-JUL-2002 (first entry)
 DE
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 3272.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 OS Homo sapiens.
 XX

PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 839 BP; 370 A; 284 C; 89 G; 96 T; 0 U; 0 Other;
 Query Match 67.7%; Score 17.6; DB 6; Length 839;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 3 AACGGCGAGCGCCGAATCTACGAA 26
 Db 382 AACGACGACGCGCCGAATCTAATAA 405
 RESULT 21
 ABQ33932/C
 ID ABQ33932 standard; DNA; 851 BP.
 XX
 AC ABQ33932;
 XX
 XX 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20523.
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX

XX 01-SEP-2001; 2001WO-EP010074.
 PF
 XX 01-SEP-2000; 2000DE-01043826.
 PR
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 851 BP; 119 A; 112 C; 357 G; 263 T; 0 U; 0 Other;
 Query Match 67.7%; Score 17.6; DB 6; Length 851;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 ACAACGGCGAGCGCCGAATCTACG 24
 Db 347 AAACGACGACGCGCCGAATCTACG 324
 RESULT 22
 ABQ33933
 ID ABQ33933 standard; DNA; 851 BP.
 XX
 AC ABQ33933;
 XX
 XX 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20524.
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX

PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. AB013410-
CC AB054121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 851 BP; 263 A; 357 C; 112 G; 119 T; 0 U; 0 Other;
SQ
Query Match 67.7%; Score 17.6; DB 6; Length 851;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGCGAGCGCCGAATCTACG 24
Db 505 AAAAAAGACGACGCCCGCAACTACG 528
RESULT 23
ACA24162
ID ACA24162 standard; DNA; 1017 BP.
XX
XX ACA24162;
XX
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #5819.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Borrelia cepacia.
XX WO20027183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX

XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU20292.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 12032; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1017 BP; 184 A; 334 C; 328 G; 171 T; 0 U; 0 Other;
SQ
Query Match 67.7%; Score 17.6; DB 7; Length 1017;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CAACGCGAGCGCCGAATCTACGA 25
Db 798 CATCGCGAGCGCCCACTACGA 821
RESULT 24
ABL16852
ID ABL16852 standard; DNA; 11340 BP.
XX
XX ABL16852;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 2029.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX

```
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 2029; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB140511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB857737-
CC AB872072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11340 BP; 3203 A; 2409 C; 2429 G; 3299 T; 0 U; 0 Other;
Query Match 67.7%; Score 17.6; DB 4; Length 11340;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGATCTTACG 24
DB 6887 ACAACGGCGCGACGATCTTACG 6910
RESULT 25
AAV30458 O/C
WP Sequence split into 6 fragments LOCUS AAV30458 Accession Aav30458
WP Fragment Name Begin End
WP AAV30458_0 1 110000
WP AAV30458_1 100001 210000
WP AAV30458_2 200001 310000
WP AAV30458_3 300001 410000
WP AAV30458_4 400001 510000
WP AAV30458_5 500001 534720
ID AAV30458 standard; DNA; 534720 BP.
XX
XX AAV30458;
XX
XX 14-OCT-1998 (first entry)
XX
XX Rhizobium species plasmid pNGR234a.
XX
XX Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW degradation; metabolism; host range; nitrogen fixation; nodulation;
KW legume; plant; ds.
XX
XX Rhizobium sp.
XX
XX Key Location/Qualifiers
FH 417796..418671
FT /*tag= a
FT /standard name= "ORF K1"
FT /product= "oligopeptide permease"
FT /note= "homologous to the OppC gene"
FT
```

```
FT 418673..419680
/*tag= b
/standard name= "ORF K2"
/product= "oligopeptide permease"
/note= "homologous to the OppD gene"
419677..420738
/*tag= c
/standard name= "ORF K3"
/product= "oligopeptide permease"
/note= "homologous to the OppF gene"
420774..422159
/*tag= d
/standard name= "ORF K4"
/product= "encapsulation-like protein"
/note= "homologous to the CapA gene"
422628..424031
/*tag= e
/standard name= "ORF K5"
/product= "aminotransferase-like protein"
/note= "homologous to the BioA gene"
424056..425594
/*tag= f
/standard name= "ORF K6"
/product= "(semi)aldehyde dehydrogenase-like protein"
complement(426949..428028)
/*tag= g
/standard name= "ORF K7"
/product= "transposase homologue"
/note= "homologous to the Trp gene"
428292..429623
/*tag= h
/standard name= "ORF K8"
/product= "glutamate dehydrogenase-like protein"
/note= "homologous to the GLUD1 gene"
complement(430538..431284)
/*tag= i
/standard name= "ORF K9"
/product= "transposase homologue"
complement(431296..432840)
/*tag= j
/standard name= "ORF K10"
/product= "transposase homologue"
/note= "homologous to the Trp gene"
complement(433880..434110)
/*tag= k
/standard name= "ORF K11"
/product= "protein of unknown function"
/note= "homologous to the FixU gene"
complement(434107..434433)
/*tag= l
/standard name= "ORF K12"
/product= "protein of unknown function"
complement(434517..434711)
/*tag= m
/standard name= "ORF K13"
/product= "ferrodoxin/ferrodoxin-like protein"
/note= "homologous to the FdxN gene"
complement(434753..436234)
/*tag= n
/standard name= "ORF K14"
/gene= "nifB"
/product= "protein involved in FeMo co-factor
biosynthesis"
complement(436460..438130)
/*tag= o
/standard name= "ORF K15"
/gene= "nifA"
/product= "positive regulator of nif, fix and other
genes"
complement(438297..438590)
/*tag= p
/standard name= "ORF K16"
/gene= "fixX"
FT
```


XX DR WPI; 1998-110606/10.

XX PT New isolated symbiotic plasmid from *Rhizobium* sp. NGR234 - used to

XX PT develop products for modifying plant characteristics, e.g. nitrogen

XX PT fixation, synthesis of compounds and stress response.

XX PS Claim 1; Fig 3; 228pp; English.

XX CC This is the nucleotide sequence of the plasmid pNGR234a isolated from

XX CC *Rhizobium* sp. NGR234. Open reading frames (ORF) derivable from the

XX CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can be

XX CC used e.g. in the transportation of compounds to and from an organism

XX CC which is a host to at least one of the nucleotide sequences, ORFs or

XX CC proteins, the degradation and/or metabolism of organic, inorganic,

XX CC natural or xenobiotic substances in a host organism or the modification

XX CC of the host range, nitrogen fixation abilities; for obtaining a synthetic

XX CC minimal set of ORFs required for functional *Rhizobium-legume* symbiosis,

XX CC especially for nodulation efficiency on host plants

XX SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 2; Length 110000;

Best Local Similarity 83.3%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGCGAGGCGCGGAATCTACGAA 26

DB 85320 AAGGCGCGCGCGGAACACGAA 85297

RESULT 27

ABQ81844

ID ABQ81844 standard; DNA; 349980 BP.

XX AC ABQ81844;

XX DT 19-NOV-2002 (first entry)

XX DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.

XX KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

XX KW antidiarrheic; antibacterial; inhibitor of *Salmonella*; detection;

XX KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

XX KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX OS Bifidobacterium longum.

XX OS Synthetic.

XX PN EP1227152-A1.

XX PD 31-JUL-2002.

XX PF 30-JAN-2001; 2001EP-00102050.

XX PR 30-JAN-2001; 2001EP-00102050.

XX PA (NEST) SOC PROD NESTLE SA.

XX XX WPI; 2002-668397/72.

XX DR Novel polynucleotide comprising Bifidobacterium genome sequence useful as

XX PT a probe or primer for detecting and/or identifying Bifidobacterium longum

XX PT in a biological sample.

XX PS Disclosure; SEQ ID NO 1100; 80pp; English.

XX CC The present invention describes a polynucleotide (I) comprising a

XX CC sequence of a Bifidobacterium genome selected from the nucleotide

XX CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

XX CC least 90% identity or which hybridises with the sequences given in

XX CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a

XX CC fusion protein, comprising a sequence selected from 1097 sequences given

CC in ABP65258 to ABP6354 ligated in frame to a polynucleotide encoding a

CC heterologous polypeptide. (I) has antidiarrheic and antibacterial

CC activities, and can be used as an inhibitor of *Salmonella*. (I) (which is

CC a probe) is useful for the detection and/or identification of

CC Bifidobacterium longum in a biological sample. A carrier containing the

CC lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM 1-2618) can be

CC used for preventing and/or treating diarrhoea brought about by pathogenic

CC bacteria and/or rotavirus. The carrier is a food composition selected

CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented

CC products, ice-creams, fermented cereal based products, milk based

CC powders, infant formula, pet food or a pharmaceutical composition

CC selected from tablets, liquid bacterial suspensions, dried oral

CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

CC (I) is useful in DNA arrays or chips to carry out analysis of the

CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent

CC Bifidobacterium related nucleotide sequences given in the Sequence

CC listing from the present invention but not mentioned further within the

CC specification. N.B. The sequence data for this patent is not represented

CC in the printed specification but is based on sequence information

CC supplied by the European Patent Office

XX SQ Sequence 349980 BP; 69975 A; 105045 C; 104394 G; 70566 T; 0 U; 0 Other;

Query Match 66.2%; Score 17.2; DB 6; Length 349980;

Best Local Similarity 86.4%; Pred. No. 3.6e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGGAGGCGCGGAATCTACGAA 26

DB 216363 CGCGGAGGCGCGGAATCTACGAA 216384

RESULT 28

ABQ14535

ID ABQ14535 standard; DNA; 530 BP.

XX AC ABQ14535;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 1126.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX KW drug; side effect; cancer; central nervous system; cardiovascular;

XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX PT WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful for

XX PT diagnosis and prognosis, comprises selective hybridization of amplicons

XX PT from chemically treated DNA.

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of

XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX CC genomic sample of DNA. The sample is treated chemically to convert